The following publication Jun Xiao, Rui Zhao, and Kin-Man Lam "Elastic net with adaptive weight for image denoising", Proc. SPIE 11515, International Workshop on Advanced Imaging Technology (IWAIT) 2020, 1151506 (1 June 2020) is available at https://doi.org/10.1117/12.2566961.

Elastic Net with Adaptive Weight for Image Denoising

Jun Xiao Rui Zhao Kin-Man Lam*
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
Hong Kong SAR, China
jun.xiao@connect.polyu.hk, rick10.zhao@connect.polyu.hk, enkmlam@polyu.edu.hk

***Abstract*—Sparse model has been widely used in image denoising, and has achieved state-of-the-art performances in past years. Dictionary learning and sparse code estimation are two key issues for the sparse model. When a dictionary is obtained, sparse code estimation is equivalent to a general least absolute shrinkage and selection operator (LASSO) problem. However, there are two limitations of LASSO: 1). LASSO gives rise to a biased estimation. 2). LASSO cannot select highly correlated features simultaneously. In recent years, methods for dictionary construction based on the non-similarity property and weighted sparse model, relying on noise estimation, have been proposed. These methods can narrow the biased gap of the estimation, and thus achieve promising results for image denoising. In this paper, we propose an elastic net with adaptive weight for image denoising. Our proposed model can achieve nearly unbiased estimation and select highly correlated features simultaneously. Experimental results show that our proposed method can obtain better performance compared with other state-of-the-art image denoising methods.**

***Index Terms*—Image denoising, weighted sparse model**

I. Introduction

Image denoising is a classic problem in low-level vision tasks, and it has been drawn researchers' attention in the past decades because of its high practical value. Noise removal is a necessary component in imaging systems, because noise is inevitably introduced and the quality of generated images is degraded. Image denoising aims to recover a latent clean image $\boldsymbol{x}$ from the observed corrupted image $\boldsymbol{y}$, i.e. $\boldsymbol{y} = \boldsymbol{x}+\boldsymbol{n}$, where $\boldsymbol{n}$ denotes the noise component and it is commonly set to the additive white Gaussian noise (AWGN).

Due to the ill-posed nature of the image-denoising problem, prior information plays an important role in denoising algorithms. The sparsity property of natural images has been proven to be useful for denoising algorithms, which assumes that the main energy of an natural image is sparsely distributed in some transformed domains, such as wavelet [1] and curvelet [2]. For the sparse model, there are two key points for image denoising: 1). The construction of a good transform domain or dictionary, and 2). The accurate estimation of the sparse code. For the dictionary construction, similar image patches are grouped to form a dictionary for image denoising in [3], which leverages the non-local similarity property. In [4], the Gaussian mixture model (GMM) is applied to image patch groups to learn their distribution, and then singular value decomposition (SVD) is employed to learn the statistical properties of the image patch group. An orthogonal dictionary is then obtained for image denoising. For learning the sparse code, a weighted sparse model [5] was proposed based on the singular matrix obtained from SVD, and achieved a better performance for image denoising. However, this reweighted strategy relies on the noise estimation techniques to update the weight iteratively. In this paper, we propose an adaptive weight strategy based on an elastic net for image denoising. Different from the traditional sparse models, our proposed method has two distinguished properties:

1) The proposed model can give rise to a nearly unbiased estimation for sparse code learning.
2) The proposed model can select two highly correlated features simultaneously, while those traditional sparse models can only select one feature.

Experimental results show that our proposed method can obtain the best performance for image denoising, compared with other state-of-the-art image denoising models.

II. SPARSE MODEL ANALYSIS

Sparse model has been well studied for image denoisinig in past years. In generally, given an degraded image $\boldsymbol{y}$ and the dictionary $\boldsymbol{D}$, the sparse model is formulated as follows:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_0, \quad (1)$$

where $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{D} \in \mathbb{R}^{p\times n}, \boldsymbol{y} \in \mathbb{R}^p$, $\hat{\boldsymbol{x}}$ is the estimated sparse code and λ controls the trade-off between the sparsity of the solution and the minimization error. Eq.(1) assumes that the dictionary $\boldsymbol{D}$ is fixed, and solving Eq.(1) is equivalent to solving the best subset selection problem, A good estimation can be obtained by existing subset selection methods, when the dimension p is small. However, when p is large, the best selection problem becomes an NP-hard problem. Instead of the L_0 norm penalty, the L_1 norm penalty is usually adopted to approximate the solution. Therefore, Eq.(1) can be rewritten as

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1. \quad (2)$$

From the view of bayesian estimation, the L_1 norm penalty can be considered as the Laplacian prior, and thus it can effectively preserve the sparsity of the solution, and obtain a better approximation of the L_0 norm penalty. Eq.(2) assumes that the dictionary is fixed, and solving the equation is equivalent to solving a general least absolute shrinkage and selection operator (LASSO) problem. In addition, a closed-form solution can be obtained by applying the soft-threshold operator, when the dictionary $\boldsymbol{D}$ is orthorgonal.

However, LASSO is a biased estimation [6], which can severely degraded the performance of some applications in the high-dimensional space, such as image denoising. In order to diminish the biased gap and further improve the performance of image denoising, an unbiased estimation and effective processing method for image denoising is necessary.

III. PROPOSED METHOD

Learning of dictionary and sparse code are the two cores in a sparse model. In this section, we will first describe how to construct the dictionary by applying the non-local similarity property, and then our proposed model will be presented. Finally, we will describe our denoising algorithm in detail.

A. Orthogonal Dictionary Design

Dictionary learning is very important for a sparse model, but online dictionary update is time consuming. Therefore, we utilize non-local similarity property of natural images to construct the dictionary.

We define a patch group (PG) as a group of similar local image patches. Given a noisy image $\boldsymbol{y}$, each image patch is extracted from a RGB image with the patch size of $p \times p \times 3$. Then, N similar local patches are searched within a region of the size $W \times W$, and each patch is stretched to form a vector $\boldsymbol{y}_n \in \mathbb{R}^{3p^2 \times 1}$ and create a PG, denoted as $\{\boldsymbol{y}_n\}_{n=1}^{N}$. We define the mean of a PG as $\boldsymbol{\mu} = \frac{1}{N}\sum_n \boldsymbol{y}_n$, and each patch in a PG is subtracted by its PG mean, which is then denoted as $\hat{\boldsymbol{Y}} = \{\hat{\boldsymbol{y}}_n = \boldsymbol{y}_n - \boldsymbol{\mu}\}$.

Assumed that L demeaned PGs are obtained, and the l-th PG is denoted as $\hat{\boldsymbol{Y}}_l = \{\hat{\boldsymbol{y}}_{l,n}\}_{n=1}^{N}$, for $i = 1, 2, \cdots, L$. In order to better describe the statistical property of each demeaned PG, SVD is applied to each $\hat{\boldsymbol{Y}}_l$, as follows:

$$\hat{\boldsymbol{Y}}_l = \boldsymbol{U}_l \boldsymbol{S}_l \boldsymbol{V}_l, \tag{3}$$

where $\boldsymbol{U}_l$ is the eigenvector matrix of the l-th demeaned mean PG, and is an orthogonal matrix. This matrix also forms the dictionary in our proposed sparse model. $\boldsymbol{S}_l$ represents singular value matrix of the l-th demeaned mean PG.

B. Elastic Net with Adaptive Weight

Consider a noisy demeaned local patch $\hat{\boldsymbol{y}}_{l,n}$, and its corresponding $\boldsymbol{U}_l$. The proposed sparse model can be formulated as

$$\begin{aligned}\hat{\boldsymbol{x}}_{l,n} = \arg\min_{\boldsymbol{x}_{l,n}} \frac{1}{2}\|\hat{\boldsymbol{y}}_{l,n} - \boldsymbol{U}_l \boldsymbol{x}_{l,n}\|^2 + \lambda\alpha\|\boldsymbol{\omega}^T \boldsymbol{x}_{l,n}\|_1 \\ + \lambda(1-\alpha)\|\boldsymbol{\omega}^T \hat{\boldsymbol{x}}_{l,n}\|^2,\end{aligned} \tag{4}$$

where $\hat{\boldsymbol{y}}_{l,n}, \boldsymbol{x}_{l,n} \in \mathbb{R}^{3p^2}, \boldsymbol{U}_l \in \mathbb{R}^{3p^2 \times 3p^2}$, and $\hat{\boldsymbol{x}}_{l,n}$ represents the estimated sparse code of the n-th local patch in the l-th PG. λ and α are tuning parameters. The L_1 norm penalty can give rise to a sparse solution, and the L_2 norm penalty promotes to group those highly correlated features. When $\alpha = 0$, Eq.(4) is equivalent to a ridge regression problem. When $\alpha = 1$, Eq.(4) is degraded to a weighted sparse problem.

In this paper, we propose a simple and effective strategy for the weight setting. We set the i-th weight for the ith element of the sparse code $\omega_i = \frac{1}{|x_i^{init}|^{\gamma}}$, where γ is a hyper parameter. The x_i^{init} can be obtained by solving a least squared regression, unless the matrix $\boldsymbol{U}_l$ is co-linear (in this case, ridge regression is preferred). From the weight setting, we can see that the initial estimation x_i^{init} contains zero components, and this will make some of the weights become infinite. We can formally handle this by introducing equality constraints into the problem of Eq.(4). In other words, those elements with infinite weights are not selected automatically when Eq.(4) is minimized. Hence, the active set of our proposed sparse model is always a subset of that of x_i^{init}. In addition, the weight setting does not rely on any noise estimation techniques, and only utilizes the information of $\hat{\boldsymbol{y}}_{l,n}$ and $\boldsymbol{U}_l$ due to the use of the least squared regression. Therefore, weights are adaptively updated according to the absolute value of the sparse code.

Since the dictionary $\boldsymbol{U}_l$ is orthogonal, Eq.(4) has a analytical solution by performing the soft-threshold operation

$$\hat{\boldsymbol{x}}_{l,n} = \frac{1}{\boldsymbol{I} + \boldsymbol{\omega}\lambda(1-\alpha)\boldsymbol{I}} \otimes S_{\boldsymbol{\omega}\lambda\alpha}(\boldsymbol{U}_l^T \hat{\boldsymbol{y}}_{l,n}), \tag{5}$$

where $\boldsymbol{\omega} = \frac{1}{|x_i^{init}|^{\gamma}}$, and $S_\lambda(\cdot)$ is the soft-threshold operator. S_λ is defined as

$$S_\lambda(\boldsymbol{x}) = \operatorname{sgn}(\boldsymbol{x}) \otimes \max(\boldsymbol{x} - \lambda, 0), \tag{6}$$

where $\operatorname{sgn}(\cdot)$ is the sign function. The threshold function of our proposed model is shown in the Fig.1. Compared with the soft-threshold function and the hard-threshold function, we can see that our proposed method can achieve nearly unbiased estimation, and eventually converge to the hard threshold function with increasing γ.

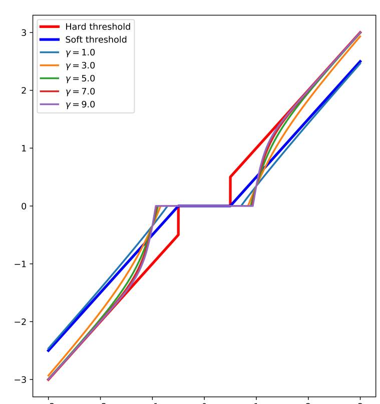

Fig. 1. Threshold function of L_0(hard), L_1(soft), and our proposed method with different γ.

C. Denoising Algorithm

When the sparse codes of PGs are obtained by solving Eq.(6), the latent clean image $\bar{\boldsymbol{y}}_{l,n}$ in $\boldsymbol{Y}_l$ is reconstructed as follows:

$$\bar{\boldsymbol{y}}_{l,n} = \boldsymbol{U}_l^T \hat{\boldsymbol{x}}_{l,n} + \boldsymbol{\mu}_l, \tag{7}$$

where $\boldsymbol{\mu}_l$ is the group mean of $\boldsymbol{Y}_l$. Then, aggregating all the reconstructed local patches to form a latent clean image. We perform the above denoising procedure, for several iterations, to obtain better denoising outputs. The proposed denoising algorithm is described in Algorithm 1.

Algorithm 1 Reweighted l_1 Norm Penalty of Sparse Model for Image Denoising

Input: Noisy image $\boldsymbol{y}$
Initialization: $\hat{\boldsymbol{x}}^{(0)} = \boldsymbol{y}$
for i=1:IteNum
1. Extracting PGs $\{\boldsymbol{Y}_n\}_{n=1}^N$
for each PG Y_n
2. Compute the group mean $\boldsymbol{\mu}_n$ and form mean subtracted PG $\bar{\boldsymbol{Y}}_n$;
3. SVD is applied to each mean subtracted PG;
4. Recover each patch in all PGs via Eq.(6) and Eq.(7)
end for
5. Aggregate the recovered PGs of all subspaces to form the recovered image $\hat{\boldsymbol{x}}^{(Ite)}$.
end for
Output: The denoised image $\hat{\boldsymbol{x}}$

IV. EXPERIMENTAL RESULTS

A. Numerical Results on Solution Paths

In this session, a simulated experiment is set up for the standard LASSO problem and our proposed method. In both models, λ is a very important regularization parameter, and different λ causes to different results of feature selection. We use 442 simulated data samples (the diabetes dataset provided by *sklearn*), and each samples has 10 dimensions. For our proposed method, γ is set to 1.0. Fig.2 plots curves of elements variance of each sparse code with respect to λ.

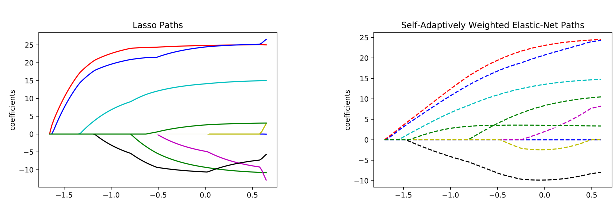

Fig. 2. The solution path of LASSO(left) and our proposed method(right).

We can see that our proposed method can active highly correlated features because of the L_2 norm penalty, while traditional LASSO does not hold this property. In other words, the blue curve and the red curve in the right side of Fig.2 have similar trend, and eventually converges as λ is increased. The green curve and the purple curve are also activated simultaneously and have similar behaviors. Detail proof about grouping correlated features of elastic net has been presented in [7].

B. Image Denoising

In this session, our proposed method is applied for image denoising. 24 color images from Kodak dataset are used in this experiment. The noisy image is synthesized by adding AWGN to R, G, and B channels respectively. Noise level set for R, G, B channels are $\sigma_r = 40, \sigma_g = 20$, and $\sigma_b = 30$. Local image patches is extracted with the size of 6 and step size of 1, and 15 similar local patches are searched within the local region of 30×30. The number of iteration is set to 2. Peak signal to noise ratio (PSNR) is adopted for assessment. We compare our proposed model with the traditional sparse model (benchmark), CBM3D [8], GMM-weighted sparse model [9] (GMM-sparse), nest image (NI), and Noise Clinic (NC) [10]. CBM3D is a classic noise removal algorithm for color images. GMM-weighted sparse model leverages external image prior to construct dictionary, and applies it to weighted sparse model for image denoising. NC is a blind image denoising method, and NI has been embedded into Photoshop and Corel PaintShop.

PSNR results of different denoising methods are listed in Table 1. The best PSNR result of each image is highlighted in bold. Compared with benchmark sparse model and GMM-sparse model, our proposed adaptive weighted elastic model greatly improves the performance, because the weighted strategy can achieve nearly unbiased estimation, and L_2 norm penalty effectively groups correlated features in the feature selection, which is a very useful property for processing high-dimensional data. In addition, our proposed method outperforms other state-of-the-art denoising models in most of testing images. Visual results of different methods are shown in the Fig.3. Compared with other denoising methods, our proposed method can effectively remove the noise while the detail of the image can be preserved.

V. CONCLUSION

Sparse model has been well studied for image denoising, but the L_1 regularized term causes to biased estimation in sparse code learning. It also cannot select highly correlated features simultaneously in high-dimensional applications, such as image denoising. In this paper, we have proposed an adaptive weighted elastic net for image denoising, whose weight updating only relies on the initial estimated value obtained from the least squared regression. Our experimental results have shown that our proposed method can achieve nearly unbiased estimation, and has the property of correlated features activation. Compared with other state-of-the-art image denoising models, our proposed method can obtain the best performance.

REFERENCES

[1] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE transactions on image processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

TABLE I
PSNR(dB) RESULTS OF DIFFERENT DENOISING METHODS ON 24 NATURAL IMAGES.

	$\sigma_r = 40, \sigma_g = 20, \sigma_b = 30$								
Image No.	Sparse	CBM3D	GMM-Sparse	NI	NC	Ours($\alpha = 0.1$)	Ours($\alpha = 0.3$)	Ours($\alpha = 0.5$)	Ours($\alpha = 0.8$)
1	24.73	25.24	25.03	23.85	24.90	25.28	25.42	**25.43**	**25.43**
2	25.68	28.27	27.36	25.90	25.87	**28.92**	28.81	28.74	28.71
3	26.14	28.81	28.05	26.00	28.58	**31.07**	30.73	30.63	30.55
4	25.97	27.95	27.65	25.82	25.67	**30.11**	29.95	29.82	29.80
5	24.70	25.03	24.84	24.38	**25.15**	24.93	25.04	25.12	25.14
6	25.27	26.24	26.03	24.65	24.74	26.78	**26.84**	26.83	26.82
7	26.06	27.88	27.33	25.63	27.69	**29.60**	29.50	29.50	29.44
8	25.03	25.05	24.96	24.02	25.30	25.46	25.62	25.67	**25.71**
9	26.04	28.44	27.64	25.94	27.44	**30.03**	29.85	29.73	29.71
10	25.83	28.27	27.55	25.87	28.42	**29.46**	29.33	28.29	29.24
11	25.12	26.95	26.05	25.32	24.67	26.58	26.59	26.59	**26.60**
12	25.87	28.76	27.63	26.01	28.37	**29.67**	29.57	29.49	29.42
13	23.35	**23.76**	23.78	23.53	22.76	22.89	23.06	23.08	23.13
14	25.02	26.02	25.53	24.94	25.68	26.05	26.09	26.12	**26.14**
15	25.85	28.38	27.18	26.06	28.21	**29.01**	28.89	28.87	28.83
16	25.67	27.75	27.10	25.69	26.66	**28.81**	28.70	28.65	28.61
17	25.65	27.90	26.89	25.85	28.32	**28.16**	28.08	28.07	28.05
18	24.90	25.77	25.66	24.74	25.70	25.93	25.99	**26.00**	25.96
19	25.83	27.30	26.78	25.40	26.52	**28.60**	28.58	28.53	28.52
20	25.62	28.96	28.05	24.95	25.90	29.43	**29.47**	29.39	29.42
21	24.14	26.54	25.94	25.40	26.52	26.61	26.63	26.68	**26.67**
22	25.53	27.05	26.87	25.36	26.60	**27.91**	27.86	27.82	27.84
23	26.18	29.14	27.70	26.13	23.24	**30.12**	29.98	29.83	29.85
24	25.28	25.75	26.16	24.55	25.73	**26.89**	26.87	27.78	27.77
Average	25.39	27.13	26.57	25.24	26.19	**27.85**	27.81	27.78	27.77

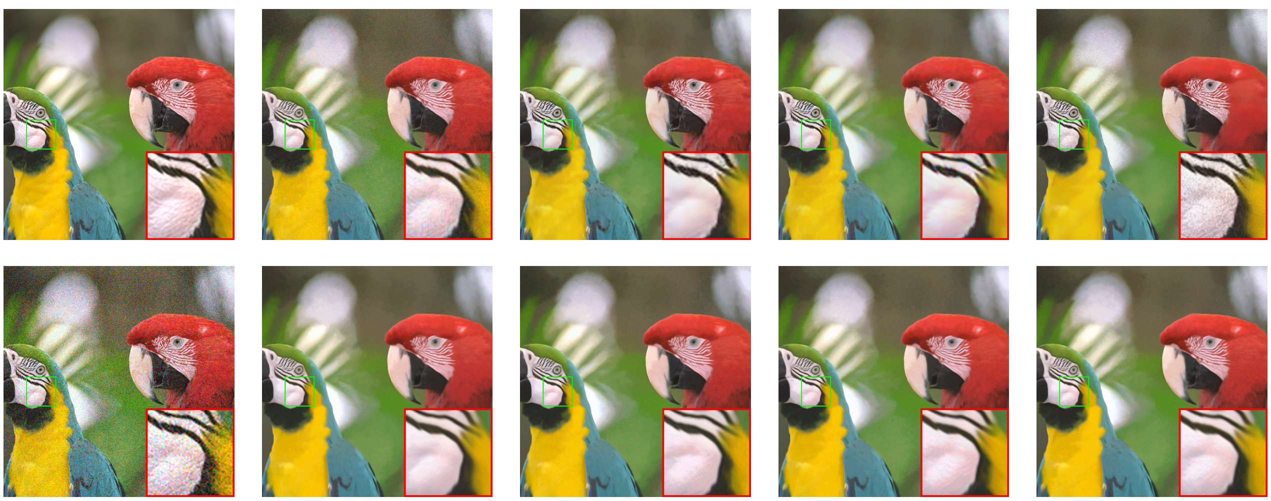

Fig. 3. Visual results of kodim 23. The first row (from left to right): ground true image, sparse model, CBM3D, GMM-sparse model, NC. The second row (from left to right): noisy image, ours($\alpha = 0.1$), ours($\alpha = 0.3$), ours($\alpha = 0.5$), ours($\alpha = 0.8$).

[2] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on image processing*, vol. 11, no. 6, pp. 670–684, 2002.

[3] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, 2014.

[4] J. Xu, L. Zhang, D. Zhang, and X. Feng, "Multi-channel weighted nuclear norm minimization for real color image denoising," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1096–1104.

[5] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng, "Patch group based nonlocal self-similarity prior learning for image denoising," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 244–252.

[6] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American statistical association*, vol. 101, no. 476, pp. 1418–1429, 2006.

[7] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[8] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space." in *ICIP (1)*, 2007, pp. 313–316.

[9] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[10] M. Lebrun, M. Colom, and J.-M. Morel, "Multiscale image blind denoising," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3149–3161, 2015.