

Multi-channel weighted nuclear norm minimization for color image denoising

Anonymous ICCV submission

Paper ID 572

Abstract

Motivated by the weighted Orthogonal Procrustes Problem, we propose a novel weighted Frobenius norm based weighted sparse coding model for non-Gaussian error modeling. We solve this model in an alternative manner. Updating of each variable has closed-form solutions and the overall model converges to a stationary point. The proposed model is applied in real image denoising problem and extensive experiments demonstrate that the proposed model can much better performance (over 1.0dB improvement on PSNR) than state-of-the-art image denoising methods, including some excellent commercial software. The novel weighted Frobenius norm can perfectly fit the non-Gaussian property of real noise.

1. Introduction

Image denoising is an important step in enhance the quality of images in computer vision systems. It aims to recover the latent clean image \mathbf{x} from the observed noisy version $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where \mathbf{n} is often assumed to be additive white Gaussian noise. Most denoising methods [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] are designed for grayscale images, and other color image denoising methods [14] treat equally the R, G, B channels in color images. However, in many computer vision tasks, the multiple channels in natural images being processed often exhibit distinct properties, e.g., contain different noise levels. For example, the noise levels among the R, G, B channels are different in real noisy images due to the on board processing in in-camera imaging pipelines [15]. This is caused by the color demosaicking during the transformation from raw data to RGB images in the standard in-camera imaging pipeline. Usually, the G channel contains the least noise levels among the three channels. Hence, in order to deal with each channel more effectively, different noise levels should be plugged into different channels for color image denoising.

The non-local self similarity (NSS) property of images has been extensively employed in image restoration tasks

such as denoising [1, 2, 3, 4, 5, 7, 8]. Among these methods, the weighted nuclear norm minimization (WNNM) model has achieved the state-of-the-art performance on denoising the additive white Gaussian noise (AWGN) in grayscale images. Though among the most effective methods, how to extend the single channel WNNM model to handle multi-channel images such as the real-world color images is still an open problem. Of course the WNNM method can be applied to denoising color images by processing each channel separately, its performance would be largely inferior than jointly processing the RGB channels by concatenating the RGB values into a single vector [14]. Besides, the searching of non-local similar patches would be unstable due to the separate processing of the RGB images and hence the power of the NSS would be largely reduced. This would also limit the performance of not only WNNM but also other NSS based methods [2, 3, 4, 5, 7]. This fact is also evaluated by our experiments on color image denoising task.

In this paper, we proposed to solve the multi-channel weighted nuclear norm minimization model to perform image denoising on color images. The original WNNM model has closed-form solutions under the weighted nuclear norm proximal operator (WNNP). However, if we add a weighting matrix \mathbf{W} to the left of the data term, the resulting multi-channel WNNM model no longer has the nice property of closed-form solutions. This makes the problem more challenging. To solve this problem, we formulate the proposed multi-channel WNNM problem into a linearly constrained non-convex program with an augmented variable. It is also not directly solvable due to the non-convexity of the existence of the weighted nuclear norm. Note that the reformulated model contains two variables with linear constraint. This can be solved by employing the alternating direction method of multipliers (ADMM) [16]. For the variable \mathbf{Z} , it is the original weighted nuclear norm minimization problem and can be solved with closed-form solution [8, 17]. For the variable \mathbf{X} , it is a standard least squares problem and we can also obtain its closed-form solution. The convergency of the proposed method is also given to guarantee a rational termination of the our model.

2. Related Work

2.1. Nuclear Norm Minimization

As the tightest convex surrogate function of the matrix rank minimization [18, 19], the nuclear norm minimization (NNM) problem has been extensively studied in low rank matrix approximation (LRMA) [20, 21, 22, 23]. A standard nuclear norm minimization problem is as follows:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_* \quad (1)$$

This NNM problem has closed-form solution by soft-thresholding the singular values of the matrix \mathbf{Y} as

$$\hat{\mathbf{X}} = \mathbf{U} \mathcal{S}_{\frac{\lambda}{2}}(\Sigma) \mathbf{V}^\top \quad (2)$$

where $\mathbf{Y} = \mathbf{U} \Sigma \mathbf{V}^\top$ is the singular value decomposition (a.k.a. Eckart-Young Decomposition [24]) of \mathbf{Y} and $\mathcal{S}_\tau(\bullet)$ is the soft-thresholding function with parameter $\tau > 0$:

$$\mathcal{S}_\tau(\Sigma_{ii}) = \max(\Sigma_{ii} - \tau, 0) \quad (3)$$

One limitation of the original NNM model is that it treats all the singular values equally but ignore the different importance of them. To make the NNM model more flexible at processing singular values, it has been extended to the truncated nuclear norm minimization model [25], the partial sum minimization of singular values [26], and the weighted nuclear norm minimization (WNNM) model [8], etc. Among these models, the WNNM model has been applied on grayscale image denoising problem with highly effective performance. This model adds weights to each singular values and the problem is:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \|\mathbf{X}\|_{w,*} \quad (4)$$

is firstly proposed for grayscale image denoising problem, where $\|\mathbf{X}\|_{w,*} = \sum_i w_i \sigma_i(\mathbf{X})$ is the weighted nuclear norm of matrix \mathbf{X} and $\mathbf{w} = [w_1, \dots, w_n]^\top$, $w_i \geq 0$ is the weight vector. According to the Remark 1 of [27], the problem (4) has closed-form solution if the weights are in a non-decreasing order

Though having achieved excellent performance on grayscale image denoising task, the WNNM method could not be applied on color image denoising in a direct manner. Of course we can apply the WNNM on each channel separately, but it has been studied that this manner would get inferior performance when compared to the power of this model on grayscale images. In this paper, we would add a weighting matrix to the WNNM model and naturally extend it to deal with color images and maintain its powerful ability on exploring the non-local self similarity property of the natural images. The proposed multi-channel WNNM model can be solved via the famous ADMM [16] algorithm. For each variable, we can derive a closed-form solution and hence the overall model can be solved in an efficient way. Besides, we analysis the convergence results of the proposed model.

2.2. Color Image Denoising

Numerous methods [1, 2, 4, 5, 7, 8, 9, 10, 11, 13] has been proposed for grayscale image denoising task. These methods could be trivially employed to denoise each channel in color images separately. For example, the CBM3D [6] first transform the RGB image into luminance-chrominance color space (e.g., YCbCr) and perform BM3D [5] for each channels separately with the patches only being grouped in luminance channel. However, as pointed out in [14], the performance would be largely reduced when compared to the performance of these methods on grayscale images. This method [14] performs color image denoising by concatenating the patches in R, G, B channels into a single vector. However, the concatenation would generate false colors and artifacts [14]. The major reason is that the concatenation methods treat multiple channels equally and ignore the different properties among these channels. To better explore the difference and correlation among the three channels in color images, several methods [28, 29, 30, 31, 32, 33, 34] have been proposed to deal with real noisy image denoising task. These methods often firstly estimate the parameters of the assumed noise model (usually Gaussian), and then perform denoising with the estimated noise model. In this paper, we propose a weighting matrix (diagonal) which add different noise levels on different channels for color image denoising. The proposed multi-channel method are able to deal with the distinct noise property among these channels. Experiments will demonstrate that the proposed model achieves better results than other competing methods for color image denoising.

3. Multi-channel Weighted Nuclear Norm Minimization

3.1. The Problem

Given In order to treat differently each channel in color images, we propose to add a weighting matrix to the original WNNM model and the resulting problem becomes

$$\min_{\mathbf{X}} \|\mathbf{W}(\mathbf{Y} - \mathbf{X})\|_F^2 + \|\mathbf{X}\|_{w,*} \quad (5)$$

where \mathbf{W} is the weighting matrix. For simplicity, we assume \mathbf{W} to be a diagonal matrix.

Unfortunately, the proposed multi-channel WNNM problem cannot be solved in an analytical form. In [27], when the weights on singular values are non-descending, the weighted nuclear norm proximal operator can have global optimum with closed-form solution. However, such property is not valid for the multi-channel WNNM model. The reason is that the weighting matrix \mathbf{W} is added to the matrix \mathbf{X} instead of its singular values. Besides, the elements in \mathbf{W} is not in a non-descending order with respect

to the singular value of \mathbf{X} . This makes the proposed model more difficult to optimize than the original WNNM model.

This can be solved by introducing an augmented variable \mathbf{Z} , and the above multi-channel WNNM problem is equivalent to a linearly constrained non-convex problem with two variables.

$$\min_{\mathbf{X}, \mathbf{Z}} \|\mathbf{W}(\mathbf{Y} - \mathbf{X})\|_F^2 + \|\mathbf{Z}\|_{w,*} \quad \text{s.t.} \quad \mathbf{X} = \mathbf{Z}. \quad (6)$$

This is an optimization problem with functions of two variables \mathbf{X} and \mathbf{Z} with linearly constrained condition of $\mathbf{X} = \mathbf{Z}$. In fact, this problem can be solved by the alternating direction method of multipliers (ADMM) algorithm, which will be introduced in the next subsection.

3.2. Optimization

To solve the above optimization problem, we first derive its augmented Lagrangian function as

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{A}, \rho) = & \|\mathbf{W}(\mathbf{Y} - \mathbf{X})\|_F^2 + \|\mathbf{Z}\|_{w,*} \\ & + \langle \mathbf{A}, \mathbf{X} - \mathbf{Z} \rangle + \frac{\rho}{2} \|\mathbf{X} - \mathbf{Z}\|_F^2 \end{aligned} \quad (7)$$

where \mathbf{A} is the augmented Lagrangian multiplier and $\rho > 0$ is the penalty parameter. After some simple calculations, we can obtain the following equivalent form of the Lagrangian function

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{A}, \rho) = & \|\mathbf{W}(\mathbf{Y} - \mathbf{X})\|_F^2 + \|\mathbf{Z}\|_{w,*} \\ & + \frac{\rho}{2} \|\mathbf{X} - \mathbf{Z} + \rho^{-1} \mathbf{A}\|_F^2 \end{aligned} \quad (8)$$

We initialize the matrix variables \mathbf{X}_0 , \mathbf{Z}_0 , and \mathbf{A}_0 to be zero matrix of suitable size. Taking derivative of the Lagrangian function \mathcal{L} with respect to the variables \mathbf{X} and \mathbf{Z} and setting the derivative function to be zero, we can alternatively update the iterations of the ADMM algorithm as follows:

(1) Update \mathbf{X} while fixing \mathbf{Z} and \mathbf{A} :

$$\mathbf{X}_{k+1} = \arg \min_{\mathbf{X}} \|\mathbf{W}\mathbf{Y} - \mathbf{W}\mathbf{X}\|_F^2 + \frac{\rho}{2} \|\mathbf{X} - \mathbf{Z}_k + \rho^{-1} \mathbf{A}_k\|_F^2 \quad (9)$$

This is a mixed weighted least square and standard least square problem and we could derive its closed-form solution:

$$\mathbf{X}_{k+1} = (\mathbf{W}^2 + \frac{\rho}{2} \mathbf{I})^{-1} (\mathbf{W}^2 \mathbf{Y} + \frac{\rho}{2} \mathbf{Z}_k - \frac{1}{2} \mathbf{A}_k) \quad (10)$$

(2) Update \mathbf{Z} while fixing \mathbf{X} and \mathbf{A} :

$$\mathbf{Z}_{k+1} = \arg \min_{\mathbf{Z}} \frac{\rho}{2} \|\mathbf{Z} - (\mathbf{X}_{k+1} + \rho^{-1} \mathbf{A}_k)\|_F^2 + \|\mathbf{Z}\|_{w,*} \quad (11)$$

According to the Theorem 1 in [27], given the $\mathbf{X}_{k+1} + \rho^{-1} \mathbf{A}_k = \mathbf{U}\Sigma\mathbf{V}^T$ be the SVD of $\mathbf{X}_{k+1} + \rho^{-1} \mathbf{A}_k$, where

Algorithm 1: Solve Multi-channel WNNM via ADMM

Input: Matrices \mathbf{Y} and \mathbf{W} , Tol > 0 , Iter = 10;

Initialization: $\mathbf{X}_0 = \mathbf{Z}_0 = \mathbf{A}_0 = \mathbf{0}$, T = False, $k = 0$;

While (T == false) **do**

1. Update \mathbf{X}_{k+1} as

$$\mathbf{X}_{k+1} = (\mathbf{W}^2 + \frac{\rho}{2} \mathbf{I})^{-1} (\mathbf{W}^2 \mathbf{Y} + \frac{\rho}{2} \mathbf{Z}_k - \frac{1}{2} \mathbf{A}_k)$$

2. Update \mathbf{Z}_{k+1} by solving the WNNM problem

$$\min_{\mathbf{Z}} \frac{\rho}{2} \|\mathbf{Z} - (\mathbf{X}_{k+1} + \rho^{-1} \mathbf{A}_k)\|_F^2 + \|\mathbf{Z}\|_{w,*}$$

3. Update \mathbf{A}_{k+1} as $\mathbf{A}_{k+1} = \mathbf{A}_k + \rho(\mathbf{X}_{k+1} - \mathbf{Z}_{k+1})$

4. $k \leftarrow k + 1$;

if ($\|\mathbf{X}_{k+1} - \mathbf{Z}_{k+1}\|_F / \|\mathbf{Z}_{k+1}\|_F < \text{Tol}$) or ($k \geq \text{Iter}$)

T \leftarrow True

end if

end while

Output: Matrices \mathbf{X} and \mathbf{Z} .

$\Sigma = \begin{pmatrix} \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times n}$, then the global optimum of the above problem is $\hat{\mathbf{Z}} = \mathbf{U}\hat{\Sigma}\mathbf{V}^T$, where $\hat{\Sigma} = \begin{pmatrix} \text{diag}(\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_n) \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times n}$ and $(\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_n)$ is the solution to the following convex optimization problem:

$$\begin{aligned} \min_{\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_n} & \sum_{i=1}^n (\sigma_i - \hat{\sigma}_i)^2 + \frac{2w_i}{\rho} \hat{\sigma}_i \\ \text{s.t.} & \hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \dots \geq \hat{\sigma}_n \geq 0. \end{aligned} \quad (12)$$

According to the Remark 1 in [27], the problem above has closed-form solution

$$\hat{\sigma}_i = \begin{cases} 0 & \text{if } c_2 < 0 \\ \frac{c_1 + \sqrt{c_2}}{2} & \text{if } c_2 \geq 0 \end{cases} \quad (13)$$

where $c_1 = \sigma_i - \epsilon$, $c_2 = (\sigma_i - \epsilon)^2 - \frac{8C}{\rho}$ and C is set as $\sqrt{2n}$ by experience in image denoising.

(3) Update \mathbf{A} while fixing \mathbf{X} and \mathbf{Z} :

$$\mathbf{A}_{k+1} = \mathbf{A}_k + \rho(\mathbf{X}_{k+1} - \mathbf{Z}_{k+1}) \quad (14)$$

We summarize the optimization steps in Algorithm 1. We can obtain a convergence theorem to guarantee that the Algorithm 1 can be terminated in a rational manner.

Theorem 1. Assume the weights in \mathbf{w} are in a non-descending order, the sequence $\{\mathbf{X}_k\}$ and $\{\mathbf{Z}_k\}$ generated in Algorithm 1 satisfy:

$$(1) \lim_{k \rightarrow \infty} \|\mathbf{X}_{k+1} - \mathbf{X}_k\| = 0; \quad (15)$$

$$(2) \lim_{k \rightarrow \infty} \|\mathbf{Z}_{k+1} - \mathbf{Z}_k\| = 0; \quad (16)$$

$$(3) \lim_{k \rightarrow \infty} \|\mathbf{X}_{k+1} - \mathbf{Z}_{k+1}\| = 0. \quad (17)$$

The proof of this theorem can be found in Appendix.

Proof. Denote

□

4. Multi-channel WNNM For Color Image Denoising

5. Experiments

5.1. Implementation Details

5.2. Experiments on Synthetic Noisy Images

5.3. Experiments on Real Noisy Images

6. Conclusion

References

- [1] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 1, 2
- [2] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. *IEEE International Conference on Computer Vision (ICCV)*, pages 2272–2279, 2009. 1, 2
- [3] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4):1620–1630, 2013. 1
- [4] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 60–65, 2005. 1, 2
- [5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 1, 2
- [6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. *IEEE International Conference on Image Processing (ICIP)*, pages 313–316, 2007. 1, 2
- [7] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng. Patch group based nonlocal self-similarity prior learning for image denoising. *IEEE International Conference on Computer Vision (ICCV)*, pages 244–252, 2015. 1, 2
- [8] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2862–2869, 2014. 1, 2
- [9] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with BM3D? *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2392–2399, 2012. 1, 2
- [10] U. Schmidt and S. Roth. Shrinkage fields for effective image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2774–2781, June 2014. 1, 2
- [11] Y. Chen, W. Yu, and T. Pock. On learning optimized reaction diffusion processes for effective image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5261–5269, 2015. 1, 2
- [12] S. Roth and M. J. Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205–229, 2009. 1
- [13] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. *IEEE International Conference on Computer Vision (ICCV)*, pages 479–486, 2011. 1, 2
- [14] Julien Mairal, Michael Elad, and Guillermo Sapiro. Sparse representation for color image restoration. *IEEE Transactions on image processing*, 17(1):53–69, 2008. 1, 2
- [15] Hakki Can Karaimer and Michael S. Brown. A software platform for manipulating the camera imaging pipeline. *European Conference on Computer Vision (ECCV)*, October 2016. 1
- [16] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011. 1, 2
- [17] Canyi Lu, Changbo Zhu, Chunyan Xu, Shuicheng Yan, and Zhouchen Lin. Generalized singular value thresholding. *AAAI*, 2015. 1
- [18] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010. 2
- [19] Maryam Fazel. Matrix rank minimization with applications. *PhD thesis, Stanford University*, 2002. 2
- [20] Nathan Srebro, Tommi Jaakkola, et al. Weighted low-rank approximations. *ICML*, 3(2003):720–727, 2003. 2
- [21] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010. 2
- [22] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011. 2
- [23] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. *Advances in neural information processing systems*, pages 612–620, 2011. 2
- [24] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. 2
- [25] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2117–2130, Sept 2013. 2

- [26] T. H. Oh, Y. W. Tai, J. C. Bazin, H. Kim, and I. S. Kweon. Partial sum minimization of singular values in robust pca: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(4):744–758, April 2016. 2
- [27] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xi-angchu Feng, and Lei Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision*, pages 1–26, 2016. 2, 3
- [28] S. Nam, Y. Hwang, Y. Matsushita, and S. J. Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1683–1691, 2016. 2
- [29] C. Liu, R. Szeliski, S. Bing Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299–314, 2008. 2
- [30] Z. Gong, Z. Shen, and K.-C. Toh. Image restoration with mixed or unknown noises. *Multiscale Modeling & Simulation*, 12(2):458–487, 2014. 2
- [31] F. Zhu, G. Chen, and P.-A. Heng. From noise modeling to blind image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 2
- [32] M. Lebrun, M. Colom, and J.-M. Morel. Multiscale image blind denoising. *IEEE Transactions on Image Processing*, 24(10):3149–3161, 2015. 2
- [33] M. Lebrun, M. Colom, and J. M. Morel. The noise clinic: a blind image denoising algorithm. <http://www.ipol.im/pub/art/2015/125/>. Accessed 01 28, 2015. 2
- [34] Neatlab ABSOft. Neat Image. <https://ni.neatvideo.com/home>. 2