IMAGE ENHANCEMENT USING TOTAL VARIATION NORM BASED CONVEX OPTIMIZATION APPROACHES

Karttikeya Mangalam (14311), Shubh Gupta (14670)

Department of Electrical Engineering, IIT Kanpur

ABSTRACT

We discuss three separate Total Variation Norm [1] based convex optimization methods for image denoising. We explain the formulation of denoising as an optimization problem and sketch the implemented algorithms for solving the problem. Also, we present several results comparing the performance of these algorithms both on perceptual quality basis and on metrics like PSNR, computation time, iterations required and rate of convergence. In conclusion, we note the utility of these algorithms in subjective parameters such as ease of implementation & parameter tuning and our remarks on the specific use-cases for each algorithm.

Index Terms— Image Denoising, Total Variation Norm, Proximal Operator, Gradient Projection

1. INTRODUCTION

Image enhancement plays an important role in many applications dealing with digital images such as television, astronomical imaging, tomography, as well as geographical information systems due to noise contamination in these systems by imperfect sensors or interference by natural phenomenon.

The objective of Image Reconstruction is to use a degraded measurement y of a high-quality image x in order to reconstruct an approximation of the image $\tilde{x}$. Degradation can be of any of the multiple forms such as blur or noise. In this project, we study and compare various convex optimization based algorithms dealing with reconstruction of a specific type of image degradation - noise. The algorithms studied model the problem of image denoising as a convex optimization problem as well as develop approximation methods to efficiently and quickly optimize it.

2. PRIOR WORK

In this section, we briefly review the various methods that have been employed in image denoising, as well as the various applications that make use of the dual decomposition technique.

One of the traditional ways employed in image denoising is by using spatial filters. Spatial filters are low pass filters applied on groups of pixels. They are based upon the idea that noise occupies the higher region in frequency spectrum of the image to be denoised, as compared to the true image. A few examples of this category are mean filter, median filter, rank-conditioned-rank-selection, as well as wiener filter [2]. A major drawback of this approach is that spatial filters remove noise to a reasonable extent but at the cost of blurring images which reduces the visibility of edges in the picture.

Filtering can also be applied in the wavelet domain, a technique proposed by Donoho [3], which gives a superior performance in image denoising due to properties such as sparsity and multi-resolution structure. Filtering in wavelet domain has also been attempted through probabilistic modeling such as Gaussian mixture models (GMMs), generalized Gaussian distributions (GGD) as well as Hidden Markov Models (HMM). Inference procedures such as Maximum A Posteriori (MAP) Estimation and Expectation-Maximization (EM) have been used in such cases to infer the parameters and coefficients of the wavelet distribution.[4]

Another approach to image denoising is to use methods which can adapt according to so give data to achieve the denoising task. Independent Component Analysis (ICA) [5] is one such method, which uses a sliding window. It requires samples from noise free data or image frames of the same scene to be denoised. An advantage of the technique is that it can denoise images with Non-Gaussian as well as Gaussian distribution of noise.

Convex Optimization has also been used to solve the denoising problem, such as Two-Step Iterative Shrinkage/ Thresholding Algorithms for Image Restoration (TwIST) [6]. Dual decomposition is another great tool in simplifying a variety of combinatorics problems which are otherwise very hard to optimize. It has many applications in a whole variety of fields such as networking problems, routing and resource allocation, control theory, machine learning as well as image deblurring and denoising tasks.

3. PROBLEM FORMULATION

Image Reconstruction is considered a highly ill posed problem in general, and traditional methods apply various constraints to model characteristics of degradation, and utilize different natural image priors to regularize the solution space.

Thanks to Prof. Ketan Rajawat for advising us.

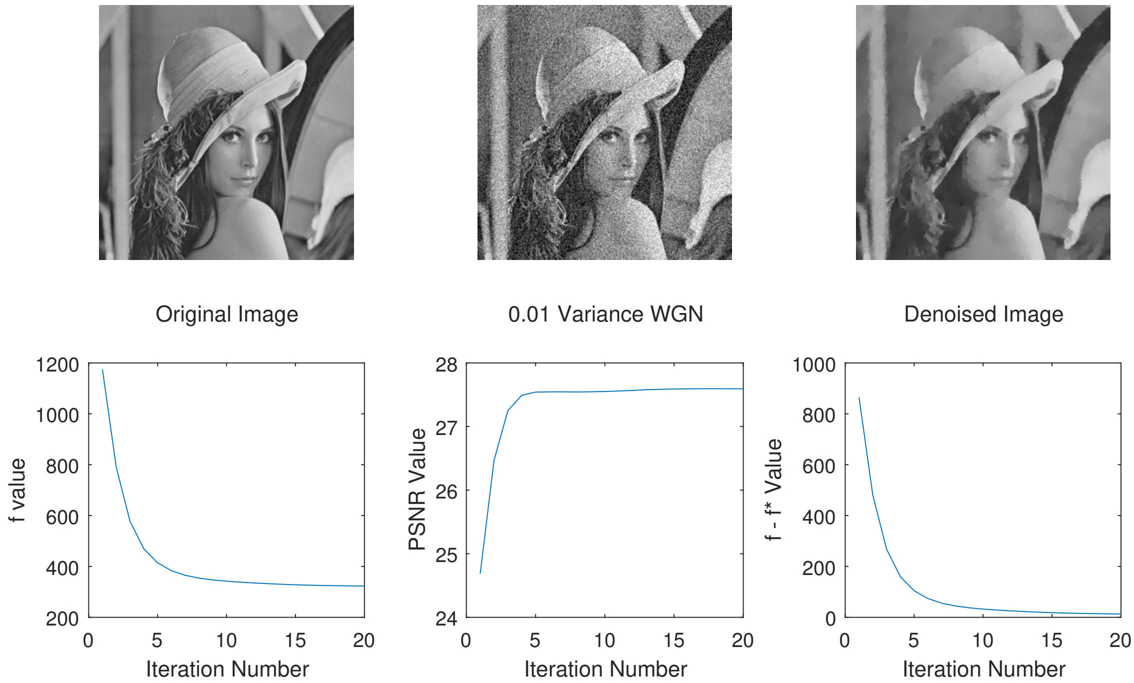

Fig. 1: a) Noisy Lena (top center) along with original (top left) and denoised image (top right) b) Number of Iteration versus Value of Optimization function (bottom left), Peak Signal to Noise Ratio (bottom center) and Accuracy (bottom right) using Fast Gradient Projection denoising method on 0.01 variance Gaussian white noise

One of the ways it is modeled usually is

$$y = DHx + \nu \tag{1}$$

where, x is the image to be restored which is not known aprior, H and D and are operators performing the degrading operations with ν as the additive noise. H and D being identities together convert the problem to a denoising problem. H being a blurring operator while D as identity makes it a deblurring problem. When D is identity and H contains a set of random projections, the problem becomes compressed sensing. When H and is a blurring operator and D is a downsampling operator it becomes a super-resolution problem. [7]

In this project, we restrict ourselves to the task of image denoising, and on approaches based on minimization of the total variation norm. The model based on total variation norm for image denoising and deblurring was proposed by Rudin-Osher and Fatemi [1] as a regularization approach for solving inverse problems. Empirical evidence has shown that it is efficient for regularizing images without blurring the sharp edges of the images. Recent researches have focused on developing very efficient algorithms for computing its proximal operator. The discrete Total Variation (TV) norm is defined as

$$||\mathbf{x}||_{TV} = \begin{cases} \sum\limits_{1\leq i,j\leq m} ||(\nabla \mathbf{x})_{i,j}||_2 & isotropic \\ \sum\limits_{1\leq i,j\leq m} |(\nabla \mathbf{x})^1_{i,j}| + |(\nabla \mathbf{x})^2_{i,j}| & anisotropic \end{cases}$$

where $(\nabla x)_{i,j}$ has components $(\nabla x)^1_{i,j}$ denoting difference with the next pixel in i direction and $(\nabla x)^2_{i,j}$ in j direction [8].

The Total Variation norm can be used to model the image denoising problem as a convex optimization problem with objective to reduce the least square error between true and noisy image with a regularization over its total variation.

$$\min_{x \in \mathcal{B}^{n\times m}_{l,u}} ||y - x||^2_F + 2\lambda ||x||_{TV} \tag{2}$$

where $\mathcal{B}^{n\times m}_{l,u}$ is an $n \times m$ matrix with box constraint over bounds (l, u) on every element.
Research in this domain has mostly been on developing algorithms to efficiently and quickly optimizing the above objective [8].

4. OUR EXPERIMENTS

In this work, we three different convex optimization algorithms for minimizing the total variation norm in the context of image denoising. These algorithms and their corresponding papers are detailed in the following subsections.

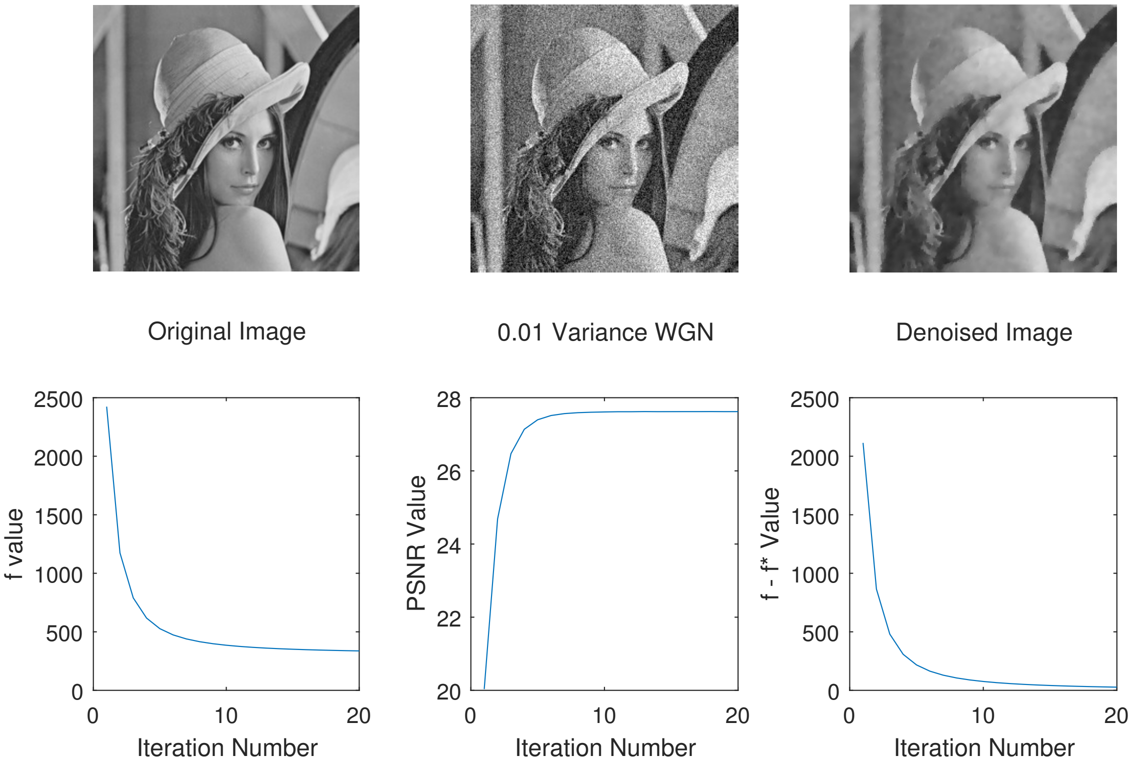

Fig. 2: a) Noisy Lena (top center) along with original (top left) and denoised image (top right) b) Number of Iteration versus Value of Optimization function (bottom left), Peak Signal to Noise Ratio (bottom center) and Accuracy (bottom right) using Gradient Projection denoising method on 0.01 variance Gaussian white noise

4.1. TV norm proximal operator

This method uses an approximation algorithm to evaluate the Total Variation proximal Operator and solve the Optimization problem (Equation 2) as proposed in [9]. Since proximal operator [10] for a function $g(x)$ is defined as

$$prox_g(\mathbf{x}) = \underset{\mathbf{y}\in\mathbf{R}^p}{\arg\min}\, g(\mathbf{y}) + (1/2)\|y - x\|_F^2$$

The solution to Equation 2 can easily found as

$$\mathbf{f}^* = \|prox_g(\mathbf{y}) - y\|_F^2 + 2\lambda\|prox_g(\mathbf{y})\|_{TV} \tag{3}$$

Since there doesn't exist a way to compute the TV norm's proximal operator; we have implemented an efficient algorithm to calculate the same based on an iterative procedure described by [9] with guaranteed convergence. Their proposed algorithm involves reformulating the proximal operator problem as a projection on a specific convex set. We are omitting the details of their procedure here in interest of space.

4.2. Gradient Projection Method

This method follows the constrained dual-based optimization algorithm proposed by Beck et al. in [8]. In this work, we have used the isotropic discrete Total Variation norm for image denoising. In [8], they have formulated the isotropic TV norm minimization as the following convex problem:

$$TV_I(\mathbf{x}) = \max_{(\mathbf{p},\mathbf{q})\in\mathcal{P}} T(\mathbf{x},\mathbf{p},\mathbf{q}) \tag{4}$$

where

$$\begin{aligned} p_{i,j} &= x_{i,j} - x_{i+1,j} \forall i \in [1, m-1], j \in [1, n] \\ q_{i,j} &= x_{i,j} - x_{i,j+1} \forall i \in [1, m], j \in [1, n-1] \end{aligned} \tag{5}$$

$$\begin{aligned} T(\mathbf{x},\mathbf{p},\mathbf{q}) = &\sum_{i=1}^{m-1}\sum_{j=1}^{n-1}[p_{i,j}(x_{i,j} - x_{i+1,j}) \\ &+ q_{i,j}(x_{i,j} - x_{i,j+1})] + \sum_{i=1}^{m-1} p_{i,n}(x_{i,n} - x_{i+1,n}) \\ &+ \sum_{j=1}^{n-1} q_{m,j}(x_{m,j} - x_{m,j+1}) \end{aligned} \tag{6}$$

Putting this formulation back in Equation 2 we obtain the following optimization problem

$$\max_{(\mathbf{p},\mathbf{q})\in P}\ \min_{\mathbf{x}\in C} \|\mathbf{x} - \mathbf{y}\|_F^2 + 2\lambda\,\mathrm{Tr}(\mathcal{L}(\mathbf{p},\mathbf{q})^{\mathbf{T}}\mathbf{x}) \tag{7}$$

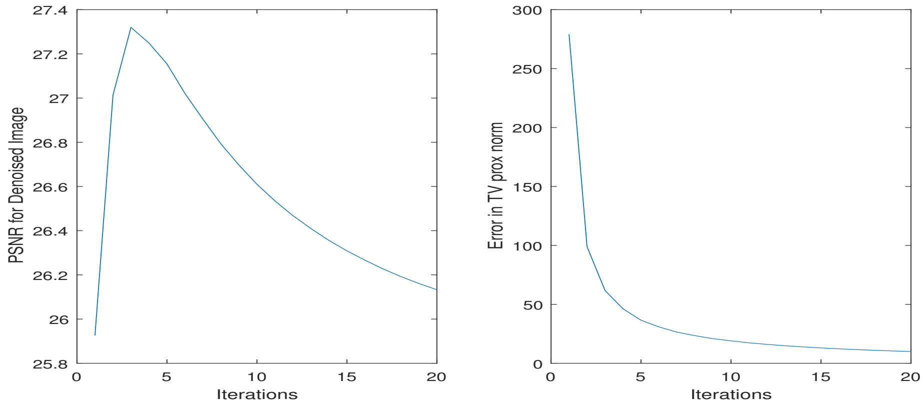

Fig. 3: Number of Iterations versus Peak Signal to Noise Ratio (left image) as well as error in TV-norm proximal operator (right image) for Total Variation norm proximal operator based denoising method on 0.01 variance Gaussian white noise

where $\mathcal{L}(\mathbf{p},\mathbf{q})$ is a linear operator constructed as

$$\begin{aligned}\mathcal{L}(\mathbf{p},\!\mathbf{q})_{i,j} &= p_{i,j} + q_{i,j} - q_{i,j-1} - p_{i-1,j} \\ &\forall\, i \in [1,m]\ j\ \in [1,n]\end{aligned} \tag{8}$$

The solution to equation (6) is then readily furnished by the following as proved in [8].

$$\mathbf{x}^* = P_C(\mathbf{y} - \lambda\mathcal{L}(\mathbf{p},\mathbf{q}))$$

with P_C being the convex orthogonal projection operator on set C. The exact denoising procedure exploiting this solution is described in [8] as **GP** algorithm.

4.3. Fast Gradient Projection Method

The basic formulation for this method is similar in spirit to GP except that it uses the Fast Iterative Shrinkage Thresholding Algorithm as introduced in the smooth function case by Nesterov [11] and extended to the discrete case in linear-inverse problem in [8]. This change gurantees a convergence rate of $O(1/k^2)$ as compared to $O(1/k)$ for gradient descent based GP algorithm. The knitty-gritty details of initialization and stopping conditions are mentioned under the algorithm name FGP in [8].

5. RESULTS & RECOMMENDATIONS

Through our experiments we observe that PSNR values obtained after image denoising are higher for FGP (27.64 dB) than those of GP (27.61 dB) or TV-norm proximal operator (21.34 dB) methods (Table 1). This implies that FGP performs similar at denoising images to GP and both perform better than

TV-norm proximal operator in terms of quality of the denoised image. Our results also show that FGP attains near-convergence values for the discrete penalized version of the unconstrained convex minimization problem for TV-based

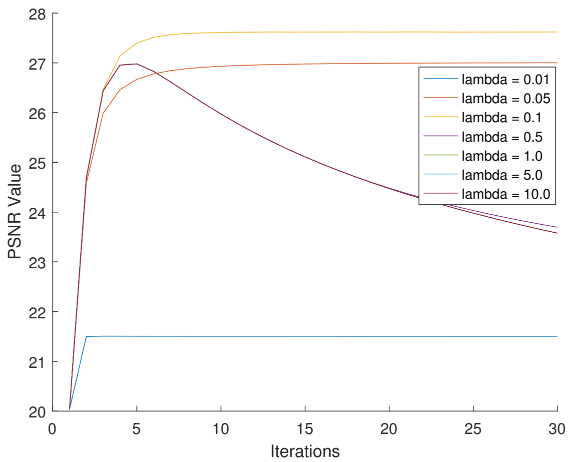

Fig. 4: Effect of Number of Iterations on Peak Signal to Noise Ratio for different λ values using Gradient Projection denoising method on 0.01 variance Gaussian white noise

deburring in fewer iterations (69) than GP (109) and TV-norm proximal operator (88) for the same image denoising task (Table 1). This is in accordance with the claim by Beck and Teboulle that FGP converges in significantly fewer iterations than the other methods, with FGP converging in $O(k^{-2})$ and GP in $O(k^{-1})$.

Another observation we make from this experiment is on the effect of changing the regularization parameter λ on the different approaches. We observe that PSNR value obtained after image denoising using FGP (Fig. 5 and 6) is quite susceptible to changes in the value of λ unlike PSNR obtained using GP (Fig. 4 and 7), which converges to the same variation over number of iterations for larger values of λ. PSNR value obtained after image denoising by TV-norm proximal operator, on the other hand, is virtually unaffected by the value of

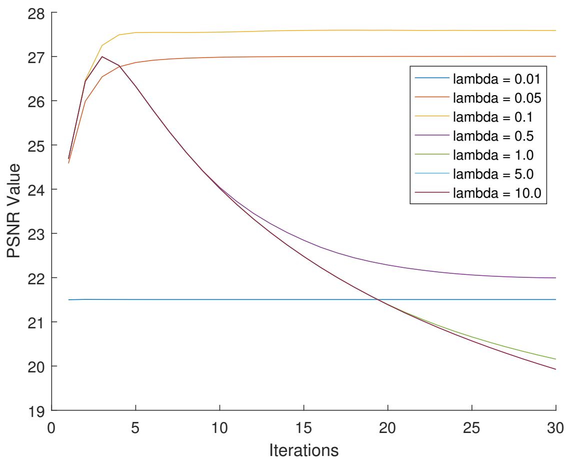

Fig. 5: Effect of Number of Iterations on Peak Signal to Noise Ratio for different λ values using Fast Gradient Projection denoising method on 0.01 variance Gaussian white noise

λ.

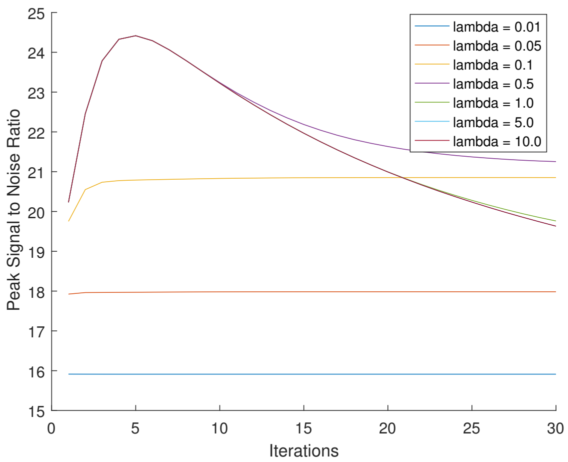

Fig. 6: Effect of Number of Iterations on Peak Signal to Noise Ratio for different λ values using Fast Gradient Projection denoising method on 0.01 variance Salt and Pepper noise

We also observed that all of the denoising methods achieve higher PSNR values in denoising gaussian noise as compared to salt and pepper noise with similar levels of image degradation (Fig. 4, 5, 6 and 7). This is because salt and pepper noise completely removes the information about the original image from the pixel unlike gaussian noise, and hence makes reconstruction harder.

It can be observed from the varying-λ plots that the approaches perform worse at denoising for lower values of lambda (Fig. 4, 5, 6 and 7), since the weight of the TV-regulariser term in optimization objective becomes insignificant for small values of λ. For $\lambda = 0$, the problem resolves to a least squares optimization problem. Another interesting observation is the effect of number of iterations on PSNR values on TV-norm proximal operator based denoised images (Fig. 3). The PSNR value first increases and then falls down unlike the PSNR obtained from GP (Fig. 2) and FGP (Fig. 1), which converge to a certain value. This demonstrates that the proximal operator for TV norm is imperfect, and in unable to model the true TV norm variation closely.

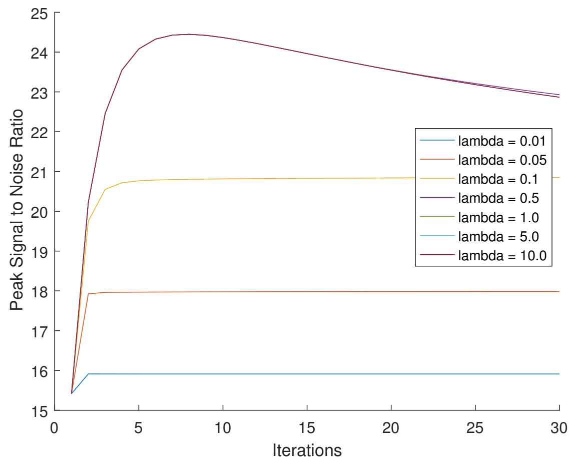

Fig. 7: Effect of Number of Iterations on Peak Signal to Noise Ratio for different λ values using Gradient Projection denoising method on 0.01 variance Salt and Pepper noise

Method	PSNR(dB)	Iterations	CPU Times (s)
TV-norm prox	27.34	88	1.23
GP	27.61	109	22.6
FGP	27.64	69	14.8

Table 1: Comparison between PSNR and convergence rates from different approaches on black and white Lena 256X256 image. Tolerance $\simeq$0.01% with an intel i7 processor for number of Iterations and CPU Times

6. CONCLUDING REMARKS

Among the algorithms for image denoising based on total variation norm minimization that we have studied, Fast Gradient Projection approach gives the most promising results, both in denoised image quality as well as convergence time. The dual based approaches - Fast Gradient Descent and Gradient Descent optimize the TV norm objective directly, and hence the PSNR values converge for large number of iterations. TV-norm proximal operator method, on the other hand, optimizes an approximation of the proximal operator of TV norm and hence the PSNR values don't converge.
Considering all that we have learnt from the experiments, the

choice of the algorithm depends on the time and quality requirement of the user - if a bit lower quality is acceptable for lesser time, TV-norm proximal operator approach would be the better choice while if quality over time is preferred, FGP is ideal. Another factor to consider is the monotonicity of the algorithm, since FGP is non-monotonic, it is harder to moniter the progress of denoising, in which case GP would be more suitable. In terms of difficulty in implementation, GP is the easiest while FGP and TV-norm proximal operator approaches are harder. Also, it is easiest to tune the hyperparameter λ for TV-norm proximal operator approach, while it is the hardest for FGP due to its high sensitivity to λ.

7. REFERENCES

[1] Leonid I Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[2] Jacob Benesty, Jingdong Chen, Yiteng Arden Huang, and Simon Doclo, "Study of the wiener filter for noise reduction," in *Speech Enhancement*, pp. 9–41. Springer, 2005.

[3] David L Donoho and Jain M Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[4] Sinisha George and Silpa Joseph, "Survey on various image denoising techniques," 2017.

[5] Andreas Jung, "An introduction to a new data analysis tool: Independent component analysis," in *Proceedings of Workshop GK" Nonlinearity"-Regensburg*, 2001.

[6] José M Bioucas-Dias and Mário AT Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image processing*, vol. 16, no. 12, pp. 2992–3004, 2007.

[7] Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838–1857, 2011.

[8] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[9] Antonin Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, no. 1-2, pp. 89–97, 2004.

[10] Neal Parikh, Stephen Boyd, et al., "Proximal algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[11] Yurii Nesterov, "A method of solving a convex programming problem with convergence rate o (1/k2)," .