Investigating convex models for image denoising

Mustafa Behrami (1009763374)
University of Melbourne
Parkville, VIC, 3010, Australia
behramim@student.unimelb.edu.au

Federico Pizarro Bejarano (1003170203)
University of Toronto
Toronto, ON M5S, Canada
federico.pizarrobejarano@mail.utoronto.ca

***Abstract*— Image denoising is an important task in computer vision. In this project, we aim to investigate algorithms that use convex optimization and are designed for denoising images corrupted by additive white Gaussian noise. Firstly, we will discuss various models of image noise. Then, we will investigate four denoising approaches: a quadratic filter, a total-variation (TV) method (solved using the primal-dual algorithm) [12], the non-local means algorithm [3], and a non-local weighted nuclear norm minimization (WNNM) filter [9]. We test these algorithms on five images corrupted by additive white Gaussian noise and and Poisson noise. We found that the non-local approaches are highly effective on additive noise but require significantly higher computation, and are largely ineffective against multiplicative noise.**

I. Introduction

Image denoising is an important task in computer vision. Imaging techniques, including ultrasound, magnetic resonance imaging (MRI), and synthetic aperture radar (SAR), are unavoidably affected by sensor noise in real applications. Recovering the original image from the observed image is an open problem in image processing.

Let us consider a standard noise model [13]

$$I_{corr} = AI + \mu, \tag{1}$$

where I_{corr} is the corrupted (noisy) image, I is the original image without any noise, A is a multiplicative noise operator, and μ is the additive noise.

Denoising images is a dynamic field that has been developing for decades, from simple smoothing, computationally expensive non-local methods, and recently machine learning [7]. Many denoising approaches are specifically tailored to the type of noise and application, such as removing the noise from MRI images which share similar structures and noise types. In this paper we will focus on the most generic type of image noise, additive zero-mean (white) Gaussian noise [7]. Specifically, we will look at algorithms that structure the denoising process as a convex optimization problem, and methods for efficiently solving these optimization problems. Through this, we gain a better understanding of convex optimization and how to apply it to varied situations.

II. Background

The noise affecting images depends on many factors, including the type of sensor, various sensor problems that may arise, and any post-processing that is done on the image. Due to this, there is no absolutely true model of image noise, but various models that have varying accuracies depending on the image in question.

This paper will not provide an extensive survey of common noise models. The denoising approaches investigated in this paper have been designed to target general noise, or specifically target zero-mean Gaussian additive noise, which is structured as

$$I_{corr} = I + \mu, \quad \mu \sim \mathcal{N}(0, \sigma^2), \tag{2}$$

where σ^2 is the variance of the Gaussian noise. However, to test the generalizability of the denoising approaches to other types of noise, we will additionally test them against Poisson noise.

Poisson noise is a multiplicative noise that represents the quantum nature of light, and thus represents a fundamental image noise that cannot be avoided [10]. Essentially, Poisson noise represents the probability of each pixel receptor getting a number of photons while taking the picture. The higher the intensity of light the pixel receptor is measuring, the higher the mean number of photons it will measure. This can be represented as

$$P(signal = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{3}$$

where k is the number of photons measured and λ is the mean number of photons it should receive [10].

III. Denoising Approaches

A. Quadratic Filtering

Quadratic filtering is a simple approach in one-dimensional signal processing that minimizes the Euclidean norm of gradient of a signal while minimizing the changes to the signal [1]:

$$\min_{\hat{x}} \|\hat{x} - x_{corr}\|_2^2 + \lambda \sum_{i}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2, \tag{4}$$

where x_{corr} is a noisy one-dimensional signal, $\hat{x}$ is the denoised signal, n is the length of the signal, and λ is a hyperparameter that defines the trade-off between smoothing the signal and maintaining the original structure. See that $\sum_{i}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2$ is the square of the euclidean norm of the gradient, $\|\nabla x_{corr}\|_2 = \sqrt{\sum_{i}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2}$. This is a convex bi-criterion optimization problem as both terms are convex [1].

Due to the simplicity of this approach and its common application to signal denoising, we applied it to image denoising as

$$\min_{\hat{I}} \|\hat{I} - I_{corr}\|_{sq} + \lambda\|\nabla_x \hat{I}\|_{sq} + \lambda\|\nabla_y \hat{I}\|_{sq}, \tag{5}$$

where $\|A\|_{sq} = \sum_{i,j} A_{ij}^2$ denotes the sum of squares norm for a matrix and ∇_x and ∇_y denote the gradients in the x and y dimensions respectively. Qualitatively, we note this approach greatly blurs the images, a known issue even in the one-dimensional case [1].

B. Total-Variation Filtering

The quadratic filter over-smooths images, resulting in blurriness. However, this can be fixed by a relatively minor change in the optimization problem. Rather than penalizing the sum-of-squares norm of the gradient, we can instead penalize the total-variation of the image . In one-dimensional signals, total variation [1][12] is simply the 1-norm of the gradient, $\sum_i^{n-1} |\hat{x}_{i+1} - \hat{x}_i|$. Similarly, we can write the quadratic denoising algorithm to use the 1-norm as

$$\min_{\hat{I}} \|\hat{I} - I_{corr}\|_{sq} + \lambda\|\hat{I}\|_{TV}, \tag{6}$$

where $\|A\|_{TV} = \sum_{i,j} \sqrt{(A_{i+1,j} - A_{i,j})^2 + (A_{i,j+1} - A_{i,j})^2}$. However, this version is difficult to minimize. A more common version is $\|A\|_{TV} = \sum_{i,j} |A_{i+1,j} - A_{i,j}| + |A_{i,j+1} - A_{i,j}|$ [5], which we see almost identical to the quadratic filter except for using an absolute value rather than square in the sum. Qualitatively, we note that this greatly decreases the blurriness of the image, but introduces a staircasing effect in the denoised images [4].

This form of total variation is more manageable, but is still more difficult to minimize than the one-dimensional case. To solve this quickly, we can make use of the primal-dual algorithm [4]. The primal-dual algorithm is a tool for minimizing optimization problems of the form:

$$\min_x F(Kx) + G(x), \tag{7}$$

where F and G are convex functions, and K is a linear operator. The convex conjugate of F is $F^*(y) = \langle Kx, y\rangle - F(Kx)$, and we can use this to rewrite the optimization as a min-max problem

$$\min_x \max_y \langle Kx, y\rangle + G(x) - F^*(Kx). \tag{8}$$

The algorithm targets this alternative formulation of the problem because it is sometimes easier to derive the proximal operator for $F^*(y)$ than it is for $F(Kx)$. The proximal operator of a convex function f is $\mathbf{prox}_f(x) = \arg\min_u (f(u) + \frac{1}{2}\|x - u\|_2^2)$, and allows for solving many convex problems. This min-max problem can be solved iteratively by improving the dual and primal problems:

$$y^{(n+1)} = \mathbf{prox}_{F^*}(y^{(n)} + \sigma K x^{(n)})$$
$$x^{(n+1)} = \mathbf{prox}_G(x^{(n)} - \tau K^* y^{(n+1)})$$

This is the Arrow-Hurwicz method, a version of the primal-dual [4]. After solving the total-variation optimization problem using this approach rather than directly using `cvxpy` [6], it was found to be approximately 330 times faster, and returned nearly identical results. This demonstrates the importance of knowing and exploiting the structure of convex optimization problems to solve them effectively and efficiently.

C. Non-Local Means

The above approaches are considered local approaches, since they look at the correlations between pixels and the pixels in their vicinity (in this case, specifically looking at the gradient). Although this works well for low noise levels, at higher noise levels the correlation between neighboring pixels is greatly perturbed [7]. Non-local self-similarity addresses this problem by considering similar patches in the image and using these repeated structures instead of local pixel correlations. This was proposed in the pioneering work of non-local means [3], where every clean pixel is the weighted average of every other pixel in the image, weighted by the similarity between the patches around the two pixels. This yields very promising image denoising results and has become the default denoiser of many applications, such as `opencv` [2]. It is not a convex optimization problem, and thus will serve as a benchmark for the convex approaches.

D. Non-Local Weighted Nuclear Norm Minimization

Another approach that combines convex optimization and NSS is non-local weighted nuclear norm minimization (WNNM) [11] [9] [8]. The nuclear norm of a matrix is simply the sum of the singular values of the matrix, which we will denote

$$\|A\|_{nuc} = \sum_i \sigma_i, \tag{9}$$

where σ_i is the i^{th} singular value of A. For small patches of an image, the singular values will describe the basic patterns in that patch. For example, a patch composed of a thick vertical line and a weaker horizontal line will have a strong and a weak singular value for the vertical and horizontal line, and smaller singular values for the (potentially noisy) flat background. Thus, to reduce the noise in the patch, we wish to minimize the small singular values while maintaining the larger singular values. This is the basic idea behind weighted nuclear norm minimization, which can be written as

$$\min_{\hat{P}} \|\hat{P} - P_{corr}\|_2^2 + \lambda\|\hat{P}\|_{w,*}, \tag{10}$$

where P_{corr} is a noisy patch of an image, $\hat{P}$ is the denoised patch, w is the set of weights to use on the singular values, and $\|A\|_{w,*} = \sum_i w_i \sigma_i$ is the weighted nuclear norm minimization which has a unique global minimum if $w_1 \geq ... \geq w_N \geq 0$ [9]. This attempts to reduce the effects of noise represented by small singular values while maintaining the structure of the image, represented by large singular values. Note that this approach cannot be applied to the entire image at once because this approach only works on simple structures

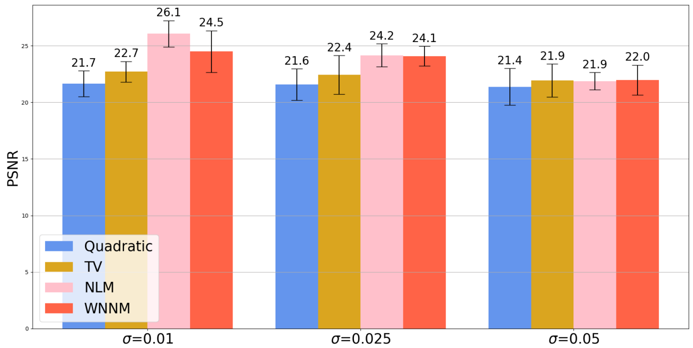

Fig. 1: Average PSNR for each algorithm on images corrupted by additive white Gaussian noise with three variances.

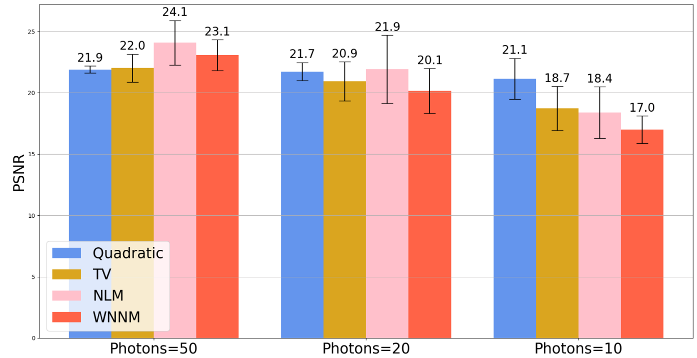

Fig. 2: Average PSNR for each algorithm on images corrupted by Poisson noise with three photon levels.

that can be represented via singular values, like lines, not full images. Additionally, the true structure of a patch can be better isolated using NSS by setting P_{corr} to be vertically stacked similar patches rather than a single patch, exploiting the repeating patterns of images. This problem has been proven to be a convex quadratic program [8].

IV. Results

A. Experimental Setup

We test the approaches on five images from the University of Southern California dataset (aerial, boat, bridge, clock, couple). Each image is scaled to 256x256 pixels and normalized to have pixel intensities between 0 and 1. The images are then corrupted by types of noise we are testing (additive white Gaussian and Poisson). After each algorithm denoises the image, the image is again normalized to have pixels between 0 and 1. Quantitative results on the denoising are done using the peak signal-to-noise ratio metric (PSNR). All results are gathered, but only some aggregated results are shown here due to space limitations. The complete results and code can be found at image_denoising.

For the non-local algorithms (NLM and WNNM), it was found to be computationally infeasible to search the entire image for similar patches. Thus, the patch side was set to be 7x7, and the search distance is 10 pixels in each direction. Additionally, WNNM requires iterating the algorithm several times, but this was not done due to computational limitations. Even with these restrictions, the non-local algorithms took several thousand times longer to run.

B. Additive White Gaussian Noise

Additive white Gaussian noise is the standard noise that these algorithms have been designed to use. Due to this, it was expected that each algorithm performs best on this type of noise. The performance of the algorithms was tested when the variance of the Gaussian noise was 0.01, 0.025, and 0.05. The results can be seen in Figure 1.

Contrary to expectations, the performance of the local methods was more robust to increasing noise-levels than the non-local approaches. In low-noise settings, NLM performed best, with WNNM closely following. However, as noise increased, all the approaches became roughly equivalent, although NLM retained the smallest variance. We believe that WNNM would outperform NLM (as claimed in the WNNM papers [9] [8]) with more computation, such as a larger search distance and iterating the algorithm. However, this could not be done due to computational limitations.

C. Poisson Noise

Poisson noise is a multiplicative noise that is based on the quantum nature of light. Rather than being additive, it depends on the intensity of each pixel and thus cannot be separated from the structure of the image. The tested algorithms were expected to suffer significantly on this type of noise. The performance of the algorithms was tested when the number of photons available per pixel was 50, 20, and 10. The results can be seen in Figure 2.

We note that in the case of a high availability of photons, the results on Poisson noise are approximately the same as those with low Gaussian noise. This is expected since the Poisson distribution approaches the Gaussian distribution at higher photon availability [10]. However, as the number of photons decrease, every approach except quadratic, especially the non-local methods, suffered greatly. Qualitatively, we find that every approach is sharper than quadratic but introduces artifacts into the image, while quadratic is consistently blurry. This shows the limitations of the PSNR metric for testing image denoising, as it is not testing the clarity of edges in the images but the overall similarity in pixel intensities.

V. Conclusions

We found that the approaches in general worked better for the Gaussian noise, as was expected, but the non-local approaches suffered at higher noise levels, which was not expected. This may be due to limiting the computation of the non-local approaches. Additionally, it was found that the naive quadratic filter was robust to changes in noise type and level, although this may be due to limitations in the PSNR metric. This paper demonstrates the importance of image denoising and the ability of convex optimization to assist in various types of denoising algorithms.

REFERENCES

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[2] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[3] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65 vol. 2, 2005.

[4] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[5] Laurent Condat. Discrete Total Variation: New Definition and Minimization. *SIAM Journal on Imaging Sciences*, 10(3):1258–1290, August 2017.

[6] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[7] Linwei Fan, Fan Zhang, Hui Fan, and Caiming Zhang. Brief review of image denoising techniques. *Visual Computing for Industry, Biomedicine, and Art*, 2(1):7, Jul 2019.

[8] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International journal of computer vision*, 121(2):183–208, 2017.

[9] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2862–2869, 2014.

[10] Samuel W. Hasinoff. *Photon, Poisson Noise*, page 608–610. Springer US, Boston, MA, 2014.

[11] 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, 1 2010.

[12] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[13] Xi-Le Zhao, Fan Wang, and Michael Ng. A new convex optimization model for multiplicative noise and blur removal. *SIAM Journal on Imaging Sciences [electronic only]*, 7, 01 2014.