An Empirical Study on Total Variation Regularization for Image Denoising

Student Name: WANG Bingbing
Student ID: 251011344R

Abstract

Image denoising is a fundamental challenge in computer vision, requiring a delicate balance between noise reduction and feature preservation. This project explores the optimization of this task using Total Variation (TV) regularization, which models denoising as a non-smooth convex optimization problem. We implement and evaluate three distinct algorithms: the Iterative Shrinkage-Thresholding Algorithm (ISTA), the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA), and the Alternating Direction Method of Multipliers (ADMM). Experiments conducted on standard benchmarks (BSDS500 and DIV2K) under varying Gaussian noise levels reveal distinct trade-offs. While ISTA demonstrates superior computational efficiency and signal fidelity (higher PSNR), ADMM exhibits remarkable robustness in preserving structural similarity (higher SSIM), particularly significantly in high-noise regimes. This study confirms the theoretical properties of these solvers and effectively highlights their practical implications in image reconstruction.

Project: https://github.com/Vincy2King/Optimization_Individual_Assignment.git

1 Introduction

In the digital era, images serve as a fundamental medium for information transmission. However, the entire pipeline of image acquisition, compression, transmission, and storage is susceptible to ubiquitous corruption by stochastic noise, such as Gaussian or salt-and-pepper impulses. This degradation extends beyond mere visual aesthetic; it critically compromises downstream high-precision computer vision tasks. For instance, noise can obscure fine-grained lesion details in medical diagnostics, generate false positives in autonomous obstacle detection, or degrade feature vectors in facial recognition systems. Consequently, image denoising is not merely a pre-processing step but a critical theoretical challenge with profound practical implications.

From a mathematical perspective, image denoising is fundamentally an ill-posed linear inverse problem. The core challenge lies in recovering a latent clean signal x from a noisy observation y, a task rife with inherent ambiguity. To resolve this, we formulate denoising as a constrained convex optimization problem. The objective is to minimize a composite energy function comprising two competing terms: a data fidelity term which ensures adherence to the observation and a regularization term which imposes prior structural knowledge, such as sparsity or smoothness.

This project specifically addresses the Total Variation (TV) regularization framework, which leverages the sparsity of image gradients to preserve sharp edges while removing noise. However, the inclusion of the L1-norm in the TV term introduces a significant optimization hurdle: non-differentiability (non-smoothness) at the origin. Standard gradient-based descent methods are inapplicable here. This necessitates the employment of sophisticated proximal algorithms and operator splitting techniques designed for non-smooth composite optimization. Bridging the gap between mathematical optimization theory and the practical challenges of image restoration, this study implements and analyzes three distinct solvers: the Iterative Shrinkage-Thresholding Algorithm (ISTA) [Daubechies et al., 2004], its momentum-based accelerated variant (FISTA) [Beck and Teboulle, 2009], and the Alternating Direction Method of Multipliers (ADMM) [Gabay and Mercier, 1976]. By benchmarking these

methods on BSDS500 [Arbelaez et al., 2010] and DIV2K [Agustsson and Timofte, 2017] datasets, we aim to dissect their convergence behaviors, computational complexity, and reconstruction fidelity. Furthermore, understanding these classical iterative solvers provides essential theoretical grounding for modern Deep Unrolling networks, effectively linking traditional optimization with advanced machine learning architectures.

2 Problem Formulation

We formulate the image denoising task as a linear inverse problem, where the objective is to recover a latent clean signal $x \in \mathbb{R}^{M \times N}$ from a noisy observation $y \in \mathbb{R}^{M \times N}$. Assuming the degradation model follows additive white Gaussian noise (AWGN), the recovery process is mathematically modeled as a minimization problem. This seeks an optimal balance between data fidelity and prior structural constraints using the following energy function:

$$\min_x \frac{1}{2}\|x - y\|_2^2 + \lambda \mathcal{R}(x) \tag{1}$$

This objective function comprises two critical components:

Data Fidelity Term: The term $(1/2) \cdot \|x - y\|_2^2$ enforces consistency with the observation. From a statistical perspective, minimizing this squared L2-norm corresponds to the maximum likelihood estimation (MLE) under the assumption of Gaussian noise. It penalizes deviations of the reconstructed image x from the noisy input y.

Regularization Term: To mitigate the ill-posed nature of the inversion, we introduce a regularization term $\lambda \mathcal{R}(x)$ that encodes prior knowledge about natural image statistics. The hyperparameter $\lambda > 0$ governs the trade-off between noise reduction and feature preservation. In this work, we employ Total Variation regularization, denoted as $\mathcal{R}(x) = \|\nabla x\|_1$. We specifically adopt the isotropic discrete TV formulation:

$$\|\nabla x\|_1 = \sum_{i,j} \sqrt{(\nabla_h x)_{i,j}^2 + (\nabla_v x)_{i,j}^2} \tag{2}$$

where ∇_h and ∇_v represent the horizontal and vertical finite difference operators, respectively. Unlike Tikhonov regularization which uses the L2-norm of gradients and leads to over-smoothing, the L1-norm in TV regularization promotes sparsity in the gradient domain. This geometric property favors piecewise-constant solutions, allowing the optimization to effectively suppress noise in smooth regions while preserving sharp discontinuities at object boundaries (edges).

Consequently, the specific optimization problem addressed in this study is formulated as:

$$\min_x \frac{1}{2}\|x - y\|_2^2 + \lambda \|\nabla x\|_1 \tag{3}$$

3 Proposed Algorithms

Due to the non-smooth and non-differentiable nature of the Total Variation term, $\|\nabla x\|_1$, in our objective function, standard gradient-based methods are not applicable. This situation provides an excellent opportunity to apply and compare a variety of advanced optimization algorithms, which directly aligns with the assignment's requirement to implement and evaluate "diverse algorithms"

3.1 Iterative Shrinkage-Thresholding Algorithm

The Iterative Shrinkage-Thresholding Algorithm (ISTA) [Daubechies et al., 2004] is a widely adopted proximal gradient method designed to solve optimization problems that involve a composite objective function. It is particularly effective for linear inverse problems where the objective function $F(x)$ consists of a smooth convex data fidelity term $f(x)$ and a non-smooth convex regularization term $g(x)$:

$$F(x) = f(x) + \lambda g(x) \tag{4}$$

where $f(x) = \frac{1}{2}\|x - y\|_2^2$ is the smooth, convex, and differentiable data fidelity term. $g(x) = \|\nabla x\|_1$ is the convex but non-smooth regularization term. The scalar λ serves as the balancing parameter. To

minimize this objective, ISTA constructs a quadratic approximation of $F(x)$ at a specific iteration point x_k. This surrogate function allows the complex optimization to be decoupled into a gradient descent step followed by a proximal projection. The iterative update rule is derived as:

$$x_{k+1} = \arg\min_x \{\frac{L}{2}\|x - (x_k - \frac{1}{L}\nabla f(x_k))\|_2^2 + \lambda g(x)\} \tag{5}$$

Here, L denotes the Lipschitz constant of the gradient $\nabla f(x)$, which controls the step size of the gradient descent to ensure convergence. For the least-squares term, the gradient is given by $\nabla f(x_k) = x_k - y$.

The iterative process can be conceptually divided into two sequential operations:

Forward Gradient Descent Step. The algorithm first updates the current estimate x_k using gradient information from the data fidelity term. This minimizes the approximation error solely with respect to the smooth component $f(x)$. We define an intermediate variable z_k as:

$$z_k = x_k - \frac{1}{L}\nabla f(x_k) \tag{6}$$

where $1/L$ acts as the step size (or learning rate). This step shifts the solution towards the noisy observation y to satisfy the data consistency constraint.

Backward Proximal Step (Shrinkage). Subsequently, the algorithm applies the regularization constraint by solving the proximal operator associated with $g(x)$. This operation maps the intermediate estimate z_k closer to the feasible set defined by the prior structure:

$$x_{k+1} = \text{prox}_{\frac{\lambda}{L}g}(z_k) \tag{7}$$

The notation $\text{prox}_{\theta g}(\cdot)$ represents the proximal operator with a parameter $\theta = \lambda/L$. Since the regularization term involves the L1-norm, the proximal operator is closely related to the Soft-Thresholding Operator, denoted as $\mathcal{S}_\theta$. For any scalar input u and threshold θ, this operator is defined as:

$$\mathcal{S}_\theta = \text{sgn}(u)\max(|u| - \theta, 0) \tag{8}$$

where $\text{sgn}(\cdot)$ is the sign function, and θ determines the shrinkage level. This operator sets small values (typically associated with noise) to zero while shrinking the magnitude of large values (associated with significant features). In the specific case of Total Variation regularization, this shrinkage effect is applied in the gradient domain, thereby suppressing spurious oscillations while preserving sharp edges in the reconstructed image x_{k+1}.

Algorithm 1 ISTA for Total Variation Denoising

Require: Noisy image y, regularization parameter λ, Lipschitz constant L, max iterations K.
Ensure: Denoised image x_K.

```
x_0 ← y, η ← 1/L                                  ▷ Initialization
for k = 0 to K − 1 do
    g_k ← x_k − y                                  ▷ Gradient of the smooth term
    z_k ← x_k − η g_k                              ▷ Forward gradient descent step
    x_{k+1} ← prox_{ηλ‖∇·‖_1}(z_k)                 ▷ Backward proximal step (TV Denoising)
end for
return x_K
```

3.2 Fast Iterative Shrinkage-Thresholding Algorithm

The Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [Beck and Teboulle, 2009] is an enhanced variant of the standard ISTA that significantly improves the convergence rate from $\mathcal{O}(1/k)$ to $\mathcal{O}(1/k^2)$. While retaining the computational simplicity of the original method, FISTA incorporates a Nesterov-style momentum term, which utilizes a linear combination of the previous two iterates to search for the optimal solution more aggressively. To implement this acceleration, the algorithm introduces an auxiliary sequence v_k that acts as a look-ahead point. Instead of performing the gradient

descent on the previous estimate x_k, the update is performed on this extrapolated point v_k. The iterative procedure consists of the following steps:

First, the auxiliary optimization variable v_k is computed as a specific linear combination of the current and previous estimates:

$$v_k = x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}) \tag{9}$$

Here, the term $x_k - x_{k-1}$ represents the inertial momentum direction, and the scalar coefficient determined by the sequence t_k controls the weight of this momentum. The scalar t_k is typically initialized as $t_1 = 1$ and updated iteratively as:

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2} \tag{10}$$

This dynamic parameter ensures that the momentum weight increases appropriately as the iterations proceed.

Second, the standard ISTA update steps, gradient descent followed by proximal mapping, are applied to the auxiliary variable v_k:

$$x_{k+1} = \text{prox}_{\frac{\lambda}{L}g}(v_k - \frac{1}{L}\nabla f(v_k)) \tag{11}$$

where $\nabla f(v_k)$ denotes the gradient of the data fidelity term evaluated at the extrapolated point v_k, and L remains the Lipschitz constant. By performing the proximal operation on this momentum-shifted point, FISTA effectively anticipates the trajectory of the optimization, leading to faster convergence towards the global minimum without increasing the computational complexity per iteration.

Algorithm 2 FISTA for Total Variation Denoising

Require: Noisy image y, parameter λ, Lipschitz constant L, max iterations K.

Ensure: Denoised image x_K.

$x_0 \leftarrow y$, $v_0 \leftarrow y$, $t_0 \leftarrow 1$, $\eta \leftarrow 1/L$ ▷ Initialization

for $k = 0$ to $K - 1$ **do**

  $g_k \leftarrow v_k - y$ ▷ Gradient at extrapolation point

  $x_{k+1} \leftarrow \text{prox}_{\eta\lambda\|\nabla\cdot\|_1}(v_k - \eta g_k)$ ▷ Proximal update

  $t_{k+1} \leftarrow (1 + \sqrt{1 + 4t_k^2})/2$ ▷ Momentum parameter update

  $v_{k+1} \leftarrow x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k)$ ▷ Auxiliary extrapolation step

end for

return x_K

3.3 Alternating Direction Method of Multipliers

The Alternating Direction Method of Multipliers (ADMM) [Gabay and Mercier, 1976] is a powerful optimization framework that combines the decomposability of dual decomposition with the robust convergence properties of the method of multipliers. It is particularly advantageous for image restoration tasks involving Total Variation, as it allows the decoupling of the non-differentiable regularization term from the data fidelity term through variable splitting.

To apply ADMM, we reformulate the original unconstrained problem into a constrained optimization problem by introducing an auxiliary variable z. We set $z = \nabla x$ to decouple the gradient operator from the image variable. The problem is rewritten as:

$$\min_{x,z} \frac{1}{2}\|x - y\|_2^2 + \lambda\|z\|_1 \quad \text{subject to} \quad z = \nabla x \tag{12}$$

Here, the variables x and z serve as the primal variables representing the image and its gradients, respectively.

The core of ADMM lies in forming the Augmented Lagrangian function. This function incorporates a quadratic penalty term to enforce the equality constraint strictly:

$$\mathcal{L}_\rho(x, z, u) = \frac{1}{1}\|x - y\|_2^2 + \lambda\|z\|_1 + \frac{\rho}{2}\|\nabla x - z + u\|_2^2 \tag{13}$$

where u represents the scaled dual variable (or Lagrange multiplier) that estimates the violation of the constraint $z = \nabla x$. The parameter $\rho > 0$ is the penalty parameter that controls the convergence speed and the rigidity of the constraint.

The optimization proceeds by minimizing $\mathcal{L}_\rho$ with respect to x and z in an alternating fashion, followed by a dual ascent update for u. This results in three sub-problems solved sequentially in each iteration:

Structural Update (x-problem) The first step minimizes the Lagrangian with respect to x while fixing z and u. Since the relevant terms are quadratic, this reduces to solving a linear system:

$$x_{k+1} = \arg\min_x (\frac{1}{2}\|x - y\|_2^2 + \frac{\rho}{2}\|\nabla x - z_k + u_k\|_2^2) \tag{14}$$

This sub-problem is differentiable and can be solved efficiently using Fast Fourier Transforms (FFT) assuming periodic boundary conditions.

Sparsity Update (z-problem) The second step minimizes with respect to z, isolating the non-differentiable L1-norm. This is solved analytically via the proximal operator:

$$z_{k+1} = \mathcal{S}_{\frac{\lambda}{\rho}}(\nabla x_{k+1} + u_k) \tag{15}$$

Here, $\mathcal{S}_{\frac{\lambda}{\rho}}$ is the Soft-Thresholding Operator defined previously. This step is responsible for enforcing the sparsity of the gradients, effectively removing noise.

Dual Update (u-update) Finally, the dual variable is updated to reflect the residual error between the gradients of the reconstructed image and the auxiliary variable:

$$u_{k+1} = u_k + \nabla x_{k+1} - z_{k+1} \tag{16}$$

Algorithm 3 ADMM for Total Variation Denoising

Require: Noisy image y, parameter λ, penalty parameter ρ, max iterations K.

Ensure: Denoised image x_K.

$x_0 \leftarrow y$, $z_0 \leftarrow \nabla y$, $u_0 \leftarrow 0$ ▷ Initialization

for $k = 0$ to $K - 1$ **do**

  $x_{k+1} \leftarrow \arg\min_x (\frac{1}{2}\|x - y\|_2^2 + \frac{\rho}{2}\|\nabla x - z_k + u_k\|_2^2)$ ▷ x-update (Structural)

  $v_{k+1} \leftarrow \nabla x_{k+1} + u_k$ ▷ Intermediate gradient computation

  $z_{k+1} \leftarrow \text{sgn}(v_{k+1}) \cdot \max(|v_{k+1}| - \lambda/\rho, 0)$ ▷ z-update (Soft-Thresholding)

  $u_{k+1} \leftarrow u_k + \nabla x_{k+1} - z_{k+1}$ ▷ u-update (Dual ascent)

end for

return x_K

4 Experiment

In this section, we evaluate the performance of three classic optimization algorithms: ISTA, FISTA, and ADMM, in solving the Total Variation regularized image denoising problem: $\min_x \frac{1}{2}\|x - y\|_2^2 + \lambda\mathcal{R}(x)$. The experiments are designed to provide a comparative analysis of their reconstruction quality, convergence dynamics, and computational efficiency.

4.1 Datasets and Metrics

Datasets. To ensure the robustness of our evaluation, we utilize two distinct large-scale datasets: the Berkeley Segmentation Dataset and Benchmark (BSDS500) [Arbelaez et al., 2010] and the DIVerse 2K resolution image dataset (DIV2K) [Agustsson and Timofte, 2017]. The BSDS500 dataset provides a standard benchmark for natural image processing, while the DIV2K dataset offers high-resolution images rich in texture and detail. For the DIV2K dataset, images exceeding 512 pixels are scaled to balance computational load with image quality.

Metrics. We employ Peak Signal-to-Noise Ratio (PSNR) [Chan and Whiteman, 1983] and Structural Similarity Index (SSIM) [Wang et al., 2004] as the primary metrics for quantitative quality assessment, alongside running time (measured in seconds) to evaluate computational efficiency.

4.2 Experimental Setting

The degradation model assumes Additive Gaussian White Noise, formulated as $y = x + n$, where $n \sim N(0, \sigma^2)$. We simulate three noise intensity levels: light ($\sigma = 15$), medium ($\sigma = 25$), and heavy ($\sigma = 50$). For the optimization parameters, the regularization weight is fixed at $\lambda = 0.1$ for all methods. In the ADMM implementation, the penalty parameter is set to $\rho = 1.0$. All algorithms are configured with a maximum of 100 iterations and a convergence threshold of 10^{-6} for the relative change in the objective function value.

4.3 Quantitative Results

The quantitative performance, summarized in Table 1, reveals a clear and consistent trade-off between pixel-level accuracy and structural preservation across the three algorithms.

Fidelity (PSNR): In terms of pixel-wise fidelity, ISTA consistently achieves the highest PSNR scores across both datasets and all noise levels. This result is expected, as ISTA is a proximal gradient method that directly minimizes an objective containing the mean-squared-error term, which PSNR is logarithmically based on. For instance, on the DIV2K dataset with $\sigma = 15$, ISTA (19.16 dB) significantly outperforms ADMM (11.25 dB).

Structure (SSIM): Conversely, for structural preservation, ADMM demonstrates remarkable robustness, particularly in medium-to-high noise regimes ($\sigma = 25$ and $\sigma = 50$). Its superiority in SSIM is attributable to the variable splitting mechanism, which decouples the gradient operator from the data fidelity term. This allows the algorithm to more effectively preserve geometric structures and textures. At $\sigma = 25$, ADMM attains the highest SSIM scores on both BSDS500 (44.81%) and DIV2K (47.00%), significantly outperforming the gradient-based methods.

Table 1: Comparison of Average PSNR (dB) and SSIM (%) results on BSDS500 and DIV2K datasets across different noise levels (σ).

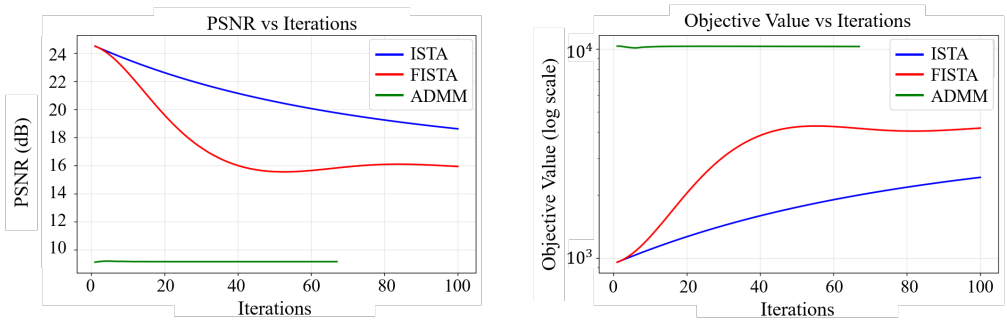
Algorithm	Average PSNR (dB)			Average SSIM (%)		
	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
BSDS500 Dataset						
ISTA	**18.86**	**16.51**	**12.80**	**33.90**	25.54	14.63
FISTA	16.13	14.61	11.78	25.87	20.87	12.86
ADMM	11.29	11.32	10.73	33.76	**44.81**	**18.97**
DIV2K Dataset						
ISTA	**19.16**	**16.72**	**12.94**	**49.00**	37.76	21.90
FISTA	16.37	14.80	11.92	38.34	31.04	19.11
ADMM	11.25	11.22	10.68	45.64	**47.00**	**26.40**

4.4 Convergence Analysis

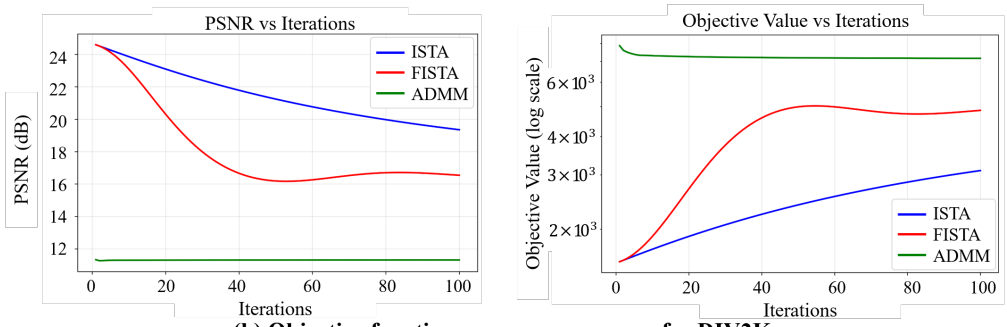
To understand the optimization dynamics, we analyze the objective function values against iterations, as shown in Figure 1. The convergence behaviors are consistent across both datasets. FISTA demonstrates the fastest descent rate during the initial 20 iterations, validating the acceleration effect of the momentum term. However, ISTA, despite a slower initial convergence, eventually reaches a lower objective function value, correlating with its higher final PSNR. In contrast, the convergence curve of ADMM is flatter and slower to minimize the primal objective compared to the gradient-based methods. The curves for the DIV2K dataset appear smoother compared to those for the BSDS500 dataset, reflecting the stability of optimization on higher-quality data.

4.5 Efficiency Analysis

We evaluate computational efficiency by comparing the running time required to meet the stopping criterion. As illustrated in Figure 2, ISTA is the most efficient algorithm, with average running times of approximately 0.51 seconds on BSDS500 and 0.58 seconds on DIV2K. This efficiency stems from its low computational complexity per iteration $O(n)$ involving simple soft-thresholding. ADMM is the slowest method (averaging from 0.64 to 0.72 seconds) due to the overhead of updating multiple split variables (x, z, u) within the augmented Lagrangian framework. While the DIV2K images incur

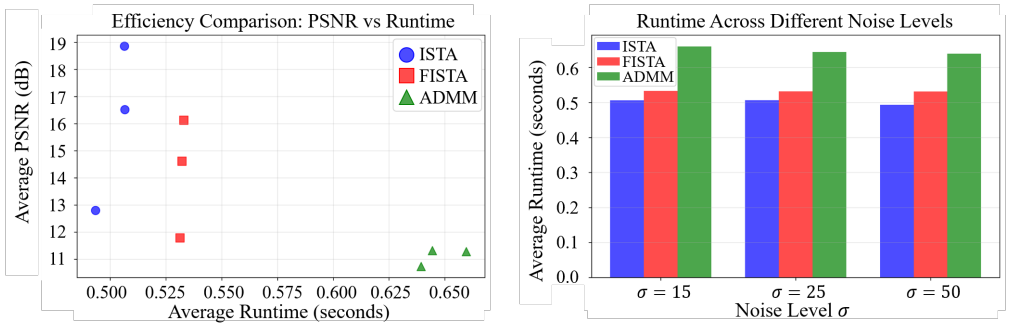


(a) Objective function convergence curves for BSDS500.

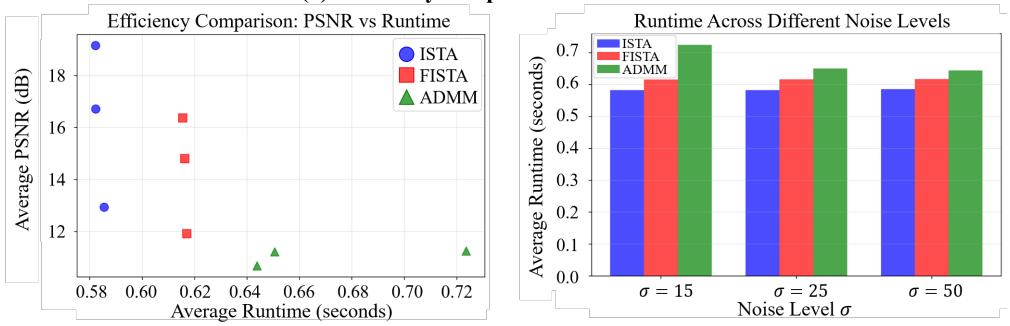


(b) Objective function convergence curves for DIV2K.

Figure 1: Objective function convergence curves. Top: BSDS500, Bottom: DIV2K.



(a) Efficiency comparison for BSDS500.



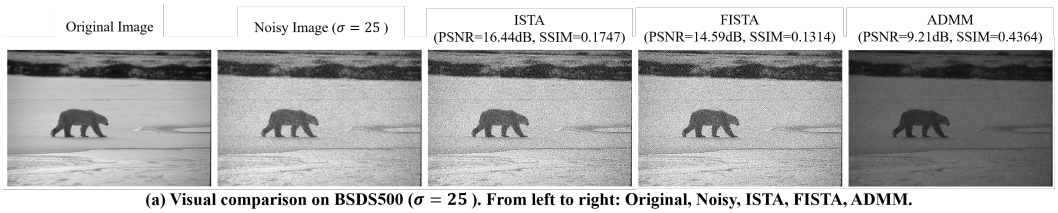
(b) Efficiency comparison for DIV2K.

Figure 2: Efficiency comparison (PSNR vs. Running Time). Top: BSDS500, Bottom: DIV2K.

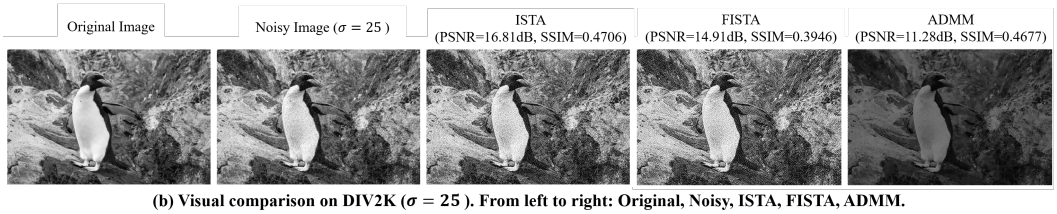
slightly higher computation times due to scaling, the linear relationship between time and image size remains consistent across datasets.

4.6 Visual Quality Assessment

Visual inspection of the denoised images confirms the quantitative metrics. Figure 3 displays the results for $\sigma = 25$ on both datasets. ISTA produces the cleanest images with sharp edges and smooth backgrounds, aligning with its high PSNR scores. FISTA yields results visually similar to ISTA but with slight over-smoothing in fine textures. Notably, despite its lower PSNR, ADMM visually retains better texture and structural details, particularly in the complex regions of the DIV2K images, preventing the "cartoon-like" artifacts sometimes seen in pure total variation minimization. This structural retention explains ADMM's superiority in SSIM scores.



(a) Visual comparison on BSDS500 ($\sigma = 25$). From left to right: Original, Noisy, ISTA, FISTA, ADMM.



(b) Visual comparison on DIV2K ($\sigma = 25$). From left to right: Original, Noisy, ISTA, FISTA, ADMM.

Figure 3: Visual comparison on BSDS500 and DIV2K ($\sigma = 25$). From left to right: Original, Noisy, ISTA, FISTA, ADMM.

Algorithm	Pros	Cons
ISTA	• **Highest Fidelity:** Consistently achieves the highest PSNR scores. • **Most Efficient:** Has the shortest running time, making it the fastest algorithm.	• **Poor Structural Preservation:** SSIM scores are significantly lower than ADMM, especially in high-noise scenarios. • **Slower Initial Convergence:** The objective function decreases more slowly than FISTA in the early iterations.
FISTA	• **Fastest Initial Convergence:** Exhibits the most rapid initial descent rate due to its momentum term.	• **Sub-optimal Final Performance:** Final PSNR and SSIM scores are not competitive with the best-performing algorithm in either metric. • **Potential for Over-smoothing:** Can cause slight over-smoothing of fine textures in visual results.
ADMM	• **Superior Structural Preservation:** Attains the highest SSIM scores, effectively preserving textures and geometric structures. • **High Visual Quality:** Avoids the "cartoon-like" artifacts often associated with TV regularization.	• **Lowest Fidelity:** Consistently yields the lowest PSNR scores. • **Least Efficient:** Is the slowest algorithm due to the overhead of updating multiple variables. • **Slow Convergence:** The primal objective function converges more slowly compared to the gradient-based methods.

Figure 4: Summary of Pros and Cons for the Evaluated Algorithms.

5 Discussion

Our empirical evaluation highlights a distinct trade-off between reconstruction fidelity, structural preservation, and computational efficiency among the different optimization algorithms, as summarized in Figure 4. For applications where raw computational speed and pixel-level accuracy (PSNR) are the primary concerns, ISTA emerges as the superior choice. It is not only the most efficient algorithm but also the most effective at minimizing the mean-squared-error. Its accelerated variant, FISTA, validates its theoretical promise by exhibiting the fastest initial convergence rate due to its momentum term, although it does not ultimately outperform ISTA in final image quality. Conversely,

when the preservation of fine textures and geometric structures is paramount, ADMM is the clear winner, consistently delivering the highest Structural Similarity Index (SSIM). This is attributable to its variable splitting mechanism, which decouples the gradient operator from the data fidelity term, thereby more effectively preserving structural information. This structural superiority, however, comes at a significant cost: ADMM is the most computationally intensive and yields the lowest PSNR. Ultimately, we can conclude that there is no universally optimal solver. The choice of algorithm is contingent upon the specific demands of the image.

6 Conclusion

This empirical study successfully integrated optimization theory with image restoration by implementing and comparing ISTA, FISTA, and ADMM for Total Variation denoising. Our results highlight that no single algorithm is universally superior; the choice depends on the specific application requirements. ISTA is the optimal choice for applications prioritizing speed and pixel-level accuracy (PSNR). Conversely, ADMM is preferable for tasks where preserving structural integrity and textural details (SSIM) is paramount, despite its higher computational cost. FISTA serves as a balanced accelerator, though it requires careful tuning in this context to match ISTA's final precision. These findings highlight the importance of selecting an optimization strategy that balances the competing demands of fidelity, structure, and efficiency.

References

Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & mathematics with applications*, 2(1):17–40, 1976.

Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):898–916, 2010.

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017.

LC Chan and P Whiteman. Hardware-constrained hybrid coding of video imagery. *IEEE Transactions on Aerospace Electronic Systems*, 19:71–84, 1983.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.