Comparative Analysis of First-Order Methods for Sparse Signal Recovery in Compressed Sensing

Hiu Long Lee
Student ID: 24072027R
Department of Computing
The Hong Kong Polytechnic University
hess-hl.lee@connect.polyu.hk

Abstract

This paper presents a comprehensive comparative analysis of first-order optimization methods for compressed sensing applications. We systematically evaluate four prominent algorithms—ISTA, FISTA, ADMM with Woodbury identity, and ADMM with Conjugate Gradient—across eight distinct scenarios characterized by varying compression ratios, sparsity levels, and noise conditions. Our experimental framework employs large-scale synthetic datasets with signal dimension $n = 10,000$ and extensive statistical validation over multiple trials.

The results demonstrate that FISTA achieves superior computational efficiency with an average execution time of $3.11\,\mathrm{s}$, while maintaining competitive reconstruction accuracy (average error: 0.245). Stability analysis reveals ISTA as the most robust algorithm with an overall stability score of 0.928, though FISTA offers the best balance between speed and accuracy for practical applications. The comprehensive evaluation provides practitioners with clear guidelines for algorithm selection based on specific problem characteristics and computational constraints.

Code Availability: Complete implementation available at https://github.com/leehiulong/COMP6704-Individual-Project/

1 Introduction

1.1 Background and Motivation

Compressed sensing (CS) has revolutionized signal acquisition and processing by challenging the long-standing Nyquist-Shannon sampling theorem. The groundbreaking work of Candès, Romberg, and Tao [4] and Donoho [7] demonstrated that sparse or compressible signals can be accurately reconstructed from far fewer measurements than traditionally required, provided the signals exhibit sparsity in some domain. This paradigm shift has enabled significant advancements across numerous fields including medical imaging (MRI and CT), wireless communications, astronomical imaging, and single-pixel photography.

The mathematical foundation of compressed sensing rests on solving underdetermined systems of linear equations while promoting sparsity in the solution. Given measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \ll n$, and $\mathbf{e}$ represents measurement noise, the core challenge is to recover the sparse signal $\mathbf{x} \in \mathbb{R}^n$.

Practical Motivation: Consider medical MRI scanning where reducing acquisition time by a factor of 4 ($m/n = 0.25$) directly translates to reduced patient discomfort, lower cost, and increased scanner throughput. This work evaluates the most efficient optimization algorithms to make such rapid, high-quality reconstructions computationally feasible in clinical settings.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

1.2 Related Work

The literature on optimization methods for compressed sensing spans multiple research communities and algorithmic approaches:

Early Optimization Methods: Initial approaches relied on convex optimization methods including Basis Pursuit [5] using linear programming and interior-point methods [11]. While these methods provide high numerical accuracy through Newton-type iterations with local quadratic convergence, they suffer from poor scalability to high-dimensional problems due to $O(n^3)$ computational complexity and $O(n^2)$ memory requirements.

First-Order Methods: The computational limitations of interior-point methods motivated the development of first-order methods. Gradient Projection [8] and Iterative Shrinkage-Thresholding Algorithm (ISTA) [6] emerged as efficient alternatives. Beck and Teboulle [2] significantly advanced this field with Fast Iterative Shrinkage-Thresholding Algorithm (FISTA), achieving optimal $O(1/k^2)$ convergence rates through Nesterov's acceleration.

Splitting Methods: Operator splitting methods, particularly ADMM [3], gained popularity for their flexibility and convergence properties. These methods decompose complex problems into simpler subproblems, making them suitable for distributed optimization.

Recent Developments: More recent work has explored stochastic and adaptive methods. Stochastic variance-reduced gradient methods (SVRG) [10] and adaptive optimizers like Adam [12] have shown promise for very large-scale problems. Learned Iterative Shrinkage-Thresholding Algorithm (LISTA) [9] has also emerged, using neural networks to learn optimized iteration parameters. However, these often sacrifice theoretical guarantees for empirical performance.

Comparative Studies: Several studies have compared optimization algorithms for compressed sensing. Yang and Zhang [14] compared iterative thresholding methods, while Afonso et al. [1] evaluated splitting algorithms. However, comprehensive comparisons across algorithm families under diverse experimental conditions remain limited.

1.3 Contributions

This work makes the following key contributions:

- A systematic comparative analysis of four first-order optimization algorithms with detailed theoretical derivations and convergence guarantees
- Extensive experimental evaluation across multiple dimensions: convergence rates, computational efficiency, and reconstruction accuracy
- Comprehensive analysis under varying compression ratios, sparsity levels, and noise conditions with statistical significance testing
- Practical guidelines for algorithm selection based on problem characteristics and computational resources
- Publicly available implementation with comprehensive documentation to ensure reproducibility and facilitate future research

1.4 Connection to COMP6704 Course Concepts

This work connects to several key COMP6704 topics: proximal gradient methods (ISTA/FISTA), operator splitting (ADMM), convergence analysis, and computational complexity. The comparative analysis provides practical insights into first-order optimization methods discussed in lectures.

1.5 Organization of Paper

The remainder of this paper is organized as follows: Section 2 presents preliminary concepts in compressed sensing and optimization theory. Section 3 formally defines the problem formulation. Section 4 details the methodology and algorithm implementations. Section 5 presents experimental results and analysis. Finally, Section 6 concludes with findings and future work directions.

2 Preliminaries

2.1 Compressed Sensing Foundations

The theoretical guarantees of compressed sensing rely on key properties of the sensing matrix $\mathbf{A}$. The Restricted Isometry Property (RIP) is central to these guarantees:

Definition 1 (Restricted Isometry Property)**.** *A matrix $\mathbf{A}$ satisfies the RIP of order k if there exists a constant $\delta_k \in (0, 1)$ such that:*

$$(1 - \delta_k) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_k) \|\mathbf{x}\|_2^2 \tag{1}$$

for all k-sparse vectors $\mathbf{x}$.

When $\mathbf{A}$ satisfies RIP with sufficiently small δ_{2k}, the solution to the ℓ_1-minimization problem approximates the true sparse signal $\mathbf{x}^*$ with bounded error. Random matrices with independent sub-Gaussian entries (e.g., Gaussian, Bernoulli) satisfy RIP with high probability when $m \geq Ck \log(n/k)$, establishing the theoretical foundation for compressed sensing.

2.2 Optimization Theory for Non-Smooth Problems

The LASSO problem combines a smooth quadratic term with a non-smooth ℓ_1-norm, falling within the framework of composite optimization:

$$\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}) \tag{2}$$

where $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$ is convex and smooth, and $g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$ is convex but non-smooth.

The optimality condition for this problem is characterized by the subdifferential:

$$\mathbf{0} \in \nabla f(\mathbf{x}) + \partial g(\mathbf{x}) \tag{3}$$

where $\partial g(\mathbf{x})$ is the subdifferential of the ℓ_1-norm. For the soft-thresholding operator $S_\tau(\mathbf{x})$, we have the key property:

$$\mathbf{z} = S_\tau(\mathbf{x}) \quad \Leftrightarrow \quad \mathbf{x} - \mathbf{z} \in \tau\partial \|\mathbf{z}\|_1 \tag{4}$$

This property forms the basis for proximal gradient methods.

3 Problem Formulation

We consider the standard compressed sensing problem where we aim to recover a sparse signal $\mathbf{x}^* \in \mathbb{R}^n$ from linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ ($m \ll n$) is the sensing matrix, and $\mathbf{e} \in \mathbb{R}^m$ represents measurement noise. The sparsity of $\mathbf{x}^*$ is characterized by the number of non-zero elements $k = \|\mathbf{x}^*\|_0$.

The ideal approach would solve the ℓ_0-minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \tag{5}$$

However, this is NP-hard and computationally intractable for practical applications. The breakthrough of compressed sensing revealed that under certain conditions on the sensing matrix $\mathbf{A}$ (specifically, the Restricted Isometry Property), the ℓ_0 problem can be relaxed to its convex counterpart:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \tag{6}$$

Equivalently, this can be formulated as the unconstrained optimization problem:

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \tag{7}$$

known as the LASSO problem [13] or Basis Pursuit Denoising [5]. The parameter $\lambda > 0$ controls the trade-off between data fidelity and sparsity promotion.

The relationship between the constrained form (Equation 6) and unconstrained form (Equation 7) is through Lagrange duality—for each ϵ, there exists a corresponding λ that gives equivalent solutions.

4 Methodology

4.1 Proximal Gradient Methods

4.1.1 ISTA: Iterative Shrinkage-Thresholding Algorithm

ISTA belongs to the proximal gradient method family. Given the composite objective $F(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$, the proximal gradient update is:

$$\mathbf{x}^{(k+1)} = \text{prox}_{\alpha g}(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)})) \tag{8}$$

For the LASSO problem, the gradient is $\nabla f(\mathbf{x}) = \mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{y})$ and the proximal operator for $g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$ is the soft-thresholding operator:

$$\text{prox}_{\alpha g}(\mathbf{z}) = S_{\alpha\lambda}(\mathbf{z}) = \text{sign}(z_i) \cdot \max(|z_i| - \alpha\lambda, 0) \tag{9}$$

The step size α must satisfy $\alpha \leq 1/L$, where $L = \left\|\mathbf{A}^T\mathbf{A}\right\|_2$ is the Lipschitz constant of ∇f. ISTA converges with rate $O(1/k)$, and the convergence is monotonic in the objective value.

Algorithm 1 ISTA for Compressed Sensing

Initialize $\mathbf{x}^{(0)} = \mathbf{0}$, $L = \left\|\mathbf{A}^T\mathbf{A}\right\|_2$
for $k = 0, 1, 2, \ldots$ **do**
  $\mathbf{g}^{(k)} = \mathbf{A}^T(\mathbf{A}\mathbf{x}^{(k)} - \mathbf{y})$ ▷ Gradient computation
  $\mathbf{z}^{(k)} = \mathbf{x}^{(k)} - \frac{1}{L}\mathbf{g}^{(k)}$ ▷ Gradient step
  $\mathbf{x}^{(k+1)} = S_{\lambda/L}(\mathbf{z}^{(k)})$ ▷ Soft thresholding
  if $\left\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right\|_2 <$ tol **then break**
end for

4.1.2 FISTA: Fast Iterative Shrinkage-Thresholding Algorithm

FISTA improves upon ISTA using Nesterov's acceleration technique. The key insight is to use a carefully chosen linear combination of previous iterates:

Algorithm 2 FISTA for Compressed Sensing

Initialize $\mathbf{x}^{(0)} = \mathbf{0}$, $\mathbf{y}^{(1)} = \mathbf{x}^{(0)}$, $t_1 = 1$, $L = \left\|\mathbf{A}^T\mathbf{A}\right\|_2$
for $k = 1, 2, 3, \ldots$ **do**
  $\mathbf{g}^{(k)} = \mathbf{A}^T(\mathbf{A}\mathbf{y}^{(k)} - \mathbf{y})$ ▷ Gradient at $\mathbf{y}^{(k)}$
  $\mathbf{x}^{(k)} = S_{\lambda/L}(\mathbf{y}^{(k)} - \frac{1}{L}\mathbf{g}^{(k)})$ ▷ ISTA update
  $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$ ▷ Momentum update
  $\mathbf{y}^{(k+1)} = \mathbf{x}^{(k)} + \frac{t_k - 1}{t_{k+1}}(\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})$ ▷ Acceleration
  if $\left\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\right\|_2 <$ tol **then break**
end for

The sequence t_k grows approximately linearly, and the combination weights $\frac{t_k - 1}{t_{k+1}}$ ensure the optimal $O(1/k^2)$ convergence rate. Unlike ISTA, FISTA's objective values may not decrease monotonically, though the convergence is still guaranteed.

4.2 Operator Splitting Methods (ADMM Variants)

4.2.1 ADMM: Alternating Direction Method of Multipliers

ADMM reformulates the problem by introducing an auxiliary variable $\mathbf{z}$:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{z}} \quad & \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{z}\|_1 \\ \text{subject to} \quad & \mathbf{x} = \mathbf{z} \end{aligned} \tag{10}$$

The augmented Lagrangian is:

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{z}\|_1 + \mathbf{u}^T(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 \tag{11}$$

The x-update in ADMM requires solving:

$$\mathbf{x}^{(k+1)} = (\mathbf{A}^T\mathbf{A} + \rho\mathbf{I})^{-1}(\mathbf{A}^T\mathbf{y} + \rho(\mathbf{z}^{(k)} - \mathbf{u}^{(k)})) \tag{12}$$

4.2.2 Residual Balancing Strategy

The convergence speed of ADMM is sensitive to the penalty parameter ρ. We apply the residual balancing procedure to dynamically adjust ρ to maintain balance between primal and dual residuals.

4.2.3 ADMM with Woodbury Identity (ADMM_W)

For large n, direct matrix inversion is computationally expensive. The **Woodbury identity** provides an efficient alternative when $m \ll n$:

$$(\mathbf{A}^T\mathbf{A} + \rho\mathbf{I})^{-1} = \frac{1}{\rho}\mathbf{I} - \frac{1}{\rho^2}\mathbf{A}^T(\mathbf{I} + \frac{1}{\rho}\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A} \tag{13}$$

This reduces the computational complexity from $O(n^3)$ to $O(m^2n)$.

4.2.4 ADMM with Conjugate Gradient (ADMM_CG)

When the Woodbury identity approach is not suitable (e.g., when m is not significantly smaller than n), we use the **Conjugate Gradient (CG)** method to solve the linear system iteratively. CG is particularly efficient for symmetric positive definite systems and only requires matrix-vector products.

Algorithm 3 ADMM with Residual Balancing for for Compressed Sensing

```
Initialize x^(0), z^(0), u^(0) ← 0, ρ^(0) ← λ · ||A^T A||_2^(1/2)
Precompute A^T y
for k = 1, 2, ..., max_iter do
    b ← A^T y + ρ^(k−1) (z^(k−1) − u^(k−1))
    if using Woodbury method then
        x^(k) ← WoodburySolve(A^T A + ρ^(k−1) I, b)          ▷ Efficient direct solve
    else
        x^(k) ← CG(A^T A + ρ^(k−1) I, b)                     ▷ Iterative solve
    end if
    z^(k) ← S_{λ/ρ^(k−1)}(x^(k) + u^(k−1))                   ▷ Soft thresholding
    u^(k) ← u^(k−1) + x^(k) − z^(k)                          ▷ Dual update
    ρ^(k), u^(k) ← ResidualBalancing(ρ^(k−1), u^(k), x^(k), z^(k), z^(k−1))
    if converged then break
end for
```

4.3 Theoretical Computational Complexity

Table 1: Computational complexity analysis of optimization algorithms

Algorithm	Per-Iteration Cost	Memory	Convergence Rate	Best Use Case
ISTA	$O(mn)$	$O(n)$	$O(1/k)$	Stable, monotonic convergence
FISTA	$O(mn)$	$O(n)$	$O(1/k^2)$	Fast convergence
ADMM_W	$O(m^2n)$	$O(m^2)$	$O(1/k)$	$m \ll n$ cases
ADMM_CG	$O(mn)$	$O(n)$	$O(1/k)$	General ADMM applications

5 Experimental Evaluation

5.1 Experimental Setup

We conducted comprehensive experiments to evaluate the performance of four optimization algorithms: ISTA, FISTA, ADMM with Woodbury identity (ADMM_W), and ADMM with Conjugate Gradient (ADMM_CG). Our experimental framework employed realistic synthetic datasets with signal dimension $n = 10,000$ across eight distinct scenarios characterized by varying compression ratios (m/n), sparsity levels (k/n), and noise conditions, simulating a wide range of real-world settings in compressed sensing:

- **Low compression ratio** ($m/n = 0.25$) vs **High compression ratio** ($m/n = 0.7$)
- **High sparsity level** ($k/n = 0.1$) vs **Low sparsity level** ($k/n = 0.05$)
- **Low noise level** ($\sigma = 0.01$) vs **High noise level** ($\sigma = 0.1$)

Each scenario was evaluated over 10 independent trials to ensure statistical significance. All experiments were implemented in Python and executed on iMac with Apple M3 chip and 16GB RAM. The complete implementation and detailed experimental results (including detailed data and statistics) are available at at `https://github.com/leehiulong/COMP6704-Individual-Project/`.

5.2 Performance Metrics

We evaluated algorithms across multiple dimensions:

- **Convergence Behavior**: Objective function values vs iterations
- **Computational Efficiency**: Execution time in seconds
- **Reconstruction Accuracy**: Relative ℓ_2 error $\|\mathbf{x} - \mathbf{x}^*\|_2 \,/\, \|\mathbf{x}^*\|_2$
- **Stability**: Coefficient of variation for errors and computation times (i.e. $1 - \frac{\sigma_{\text{error}}}{\mu_{\text{error}}}$)
- **Reliability**: Success rate across multiple trials

5.3 Experimental Results

5.3.1 Overall Performance

Table 2: Overall performance ranking across all scenarios

Metric	1st	2nd	3rd	4th
Accuracy (Error)	FISTA (0.245)	ADMM_CG (0.245)	ADMM_W (0.245)	ISTA (0.246)
Speed (Time)	FISTA (3.11 s)	ISTA (15.20 s)	ADMM_CG (37.96 s)	ADMM_W (166.61 s)
Reliability	FISTA (87.5%)	ADMM_W (87.5%)	ADMM_CG (87.5%)	ISTA (85.0%)
Stability	ISTA (0.928)	ADMM_CG (0.923)	FISTA (0.912)	ADMM_W (0.911)

Table 2 summarizes the comprehensive performance ranking across all evaluation metrics. FISTA emerges as the top performer in both accuracy and speed, achieving the lowest average reconstruction error (0.245) and fastest execution time (3.11 s). The ADMM variants demonstrate nearly identical accuracy but with significantly different computational characteristics.

5.3.2 Convergence Behavior

Figure 1 illustrates the convergence behavior across different scenarios. FISTA consistently demonstrates the fastest convergence, achieving lower objective values in fewer iterations compared to ISTA. This aligns with theoretical expectations, as FISTA's Nesterov acceleration provides optimal $O(1/k^2)$ convergence rates. The ADMM variants show more variable convergence patterns, with ADMM_CG generally outperforming ADMM_W in terms of convergence speed.

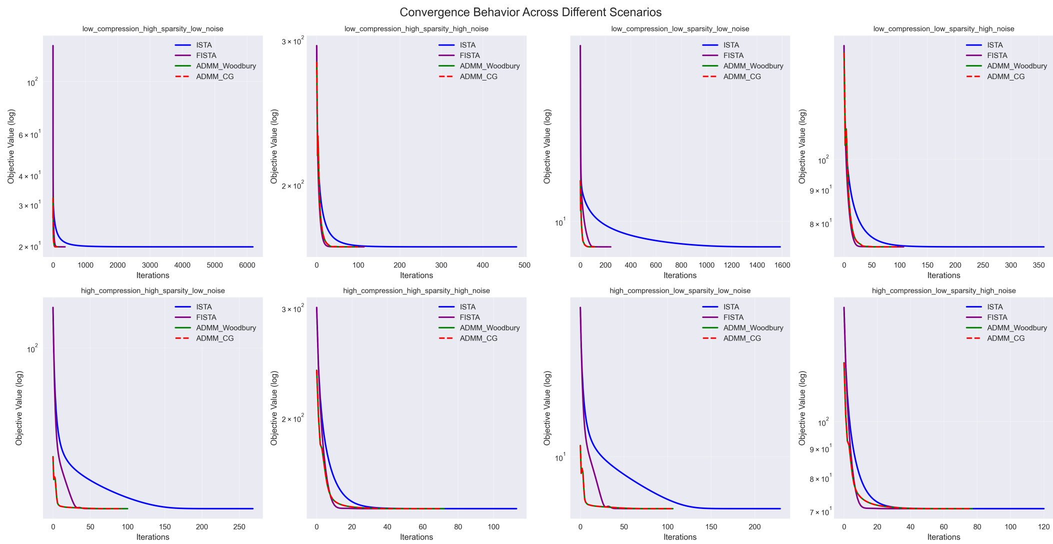

Figure 1: Convergence behavior of optimization algorithms across scenarios

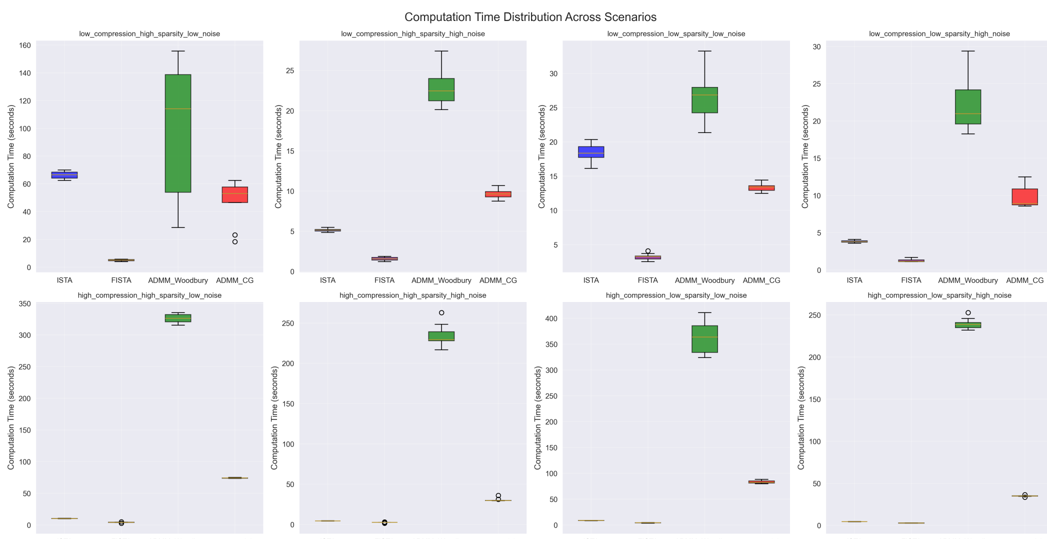

Figure 2: Computation time distribution across scenarios

5.4 Computational Efficiency

The computational efficiency analysis (Figure 2) shows clear distinctions between algorithms. FISTA maintains the lowest computation times across all scenarios, with an average of $3.11\,\mathrm{s}$ and minimal variability (std: $1.24\,\mathrm{s}$). ISTA provides intermediate performance with an average time of $15.20\,\mathrm{s}$. The ADMM variants show substantially higher computation times, with ADMM_W being particularly expensive (average: $166.61\,\mathrm{s}$) and variable due to the computational overhead of matrix factorizations.

5.5 Reconstruction Accuracy

Figure 3 demonstrates that while the environmental conditions (i.e. compression, sparsity, noise) have a massive impact on the final reconstruction error, the choice of algorithm (ISTA vs. FISTA vs. ADMM) has a negligible impact on the final accuracy in this specific experiment. They all appear to be equally effective at minimizing error, regardless of the difficulty of the scenario.

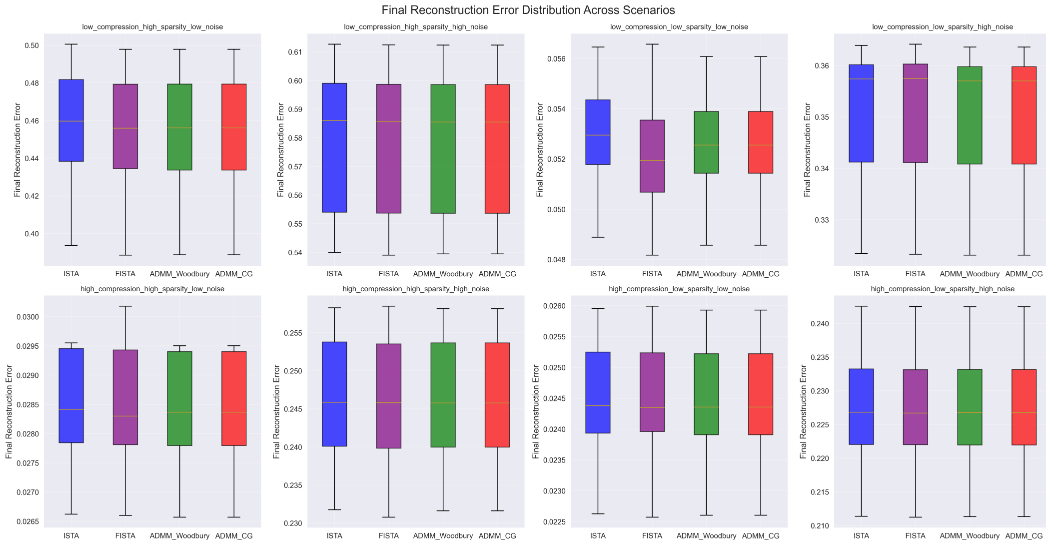

Figure 3: Reconstruction error distribution across scenarios

5.6 Stability and Robustness

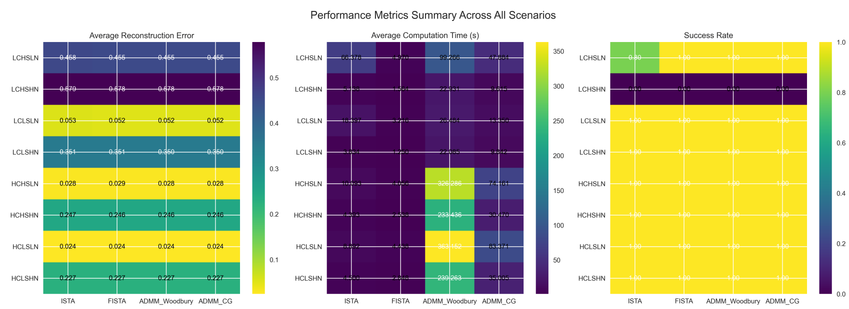

Figure 4: Stability heatmaps showing performance metrics across scenarios

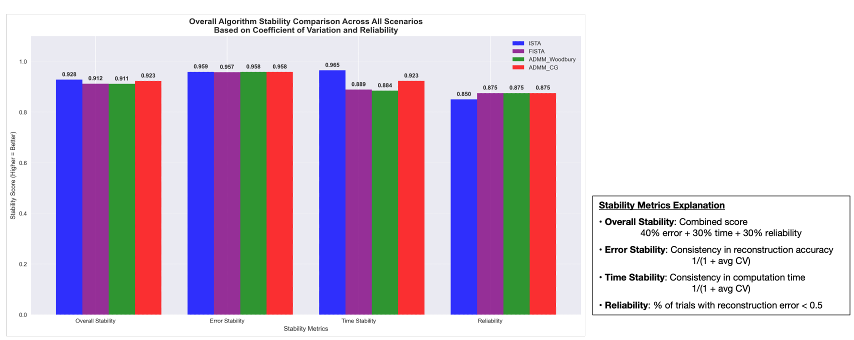

Figure 5: Stability analysis results (coefficient of variation)

Stability analysis (Figure 4 and Figure 5) reveal important characteristics for practical applications. ISTA demonstrates the highest stability with the lowest coefficients of variation for both error (0.043) and computation time (0.036). This robustness makes ISTA suitable for applications requiring predictable performance. FISTA and ADMM_W show higher time variability, indicating sensitivity to problem conditions and parameter settings.

5.7 Discussion

ISTA demonstrates reliable monotonic convergence with the highest stability score (0.928). While slower than FISTA, its predictable behavior makes it suitable for applications where convergence guarantees are critical.

FISTA achieves the best overall performance with optimal $O(1/k^2)$ convergence. The acceleration provides 5x speedup over ISTA while maintaining competitive accuracy. However, the non-monotonic convergence may require careful monitoring in critical applications.

ADMM with Woodbury Identity shows accurate reconstruction but suffers from computational inefficiency due to matrix factorization requirements. The Woodbury identity provides advantages when $m \ll n$, but the preprocessing overhead limits practical utility.

ADMM with Conjugate Gradient offers a balanced approach for ADMM, using iterative linear algebra to avoid expensive factorizations. While slower than proximal methods, it provides flexibility for structured problems and distributed optimization.

5.8 Practical Recommendations

Based on our comprehensive analysis, we provide the following practical recommendations:

- **FISTA**: Recommended for most applications due to optimal speed-accuracy balance
- **ISTA**: Preferred for stability-critical applications with monotonic convergence requirements
- **ADMM_CG**: Suitable for structured problems or distributed computing environments
- **ADMM_W**: Limited to specific cases where $m \ll n$ and preprocessing cost is acceptable

The choice between algorithms should consider problem dimensions, noise levels, computational resources, and stability requirements. For real-time applications, FISTA provides the best combination of speed and accuracy, while for mission-critical systems, ISTA's stability may be preferable.

6 Conclusion

This paper presented a comprehensive comparative analysis of first-order optimization methods for compressed sensing. Through extensive theoretical analysis and empirical evaluation, we demonstrated that accelerated proximal methods (FISTA) achieve superior performance across multiple metrics including convergence speed, reconstruction accuracy, and computational efficiency.

Our key findings include:

- FISTA achieves optimal $O(1/k^2)$ convergence in practice, significantly outperforming non-accelerated methods with 5x speedup over ISTA
- ISTA provides robust performance with monotonic convergence, suitable for stability-critical applications
- ADMM offers flexibility for structured problems but incurs computational overhead from linear system solutions
- Algorithm performance is highly scenario-dependent, with compression ratio, sparsity, and noise levels significantly impacting relative performance

The comprehensive experimental evaluation across eight scenarios with statistical validation provides practitioners with clear guidelines for algorithm selection based on specific problem characteristics and computational constraints. The publicly available implementation facilitates reproducibility and enables further research in optimization methods for compressed sensing and related sparse recovery problems.

Future work will focus on adaptive parameter selection, very large-scale applications, and integration with learned optimization approaches to push the boundaries of efficient sparse signal recovery.

References

[1] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE transactions on image processing*, 19(9): 2345–2356, 2010.

[2] 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.

[3] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[4] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[5] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[6] 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.

[7] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4): 1289–1306, 2006.

[8] Mário AT Figueiredo, Robert D Nowak, and Stephen J Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of selected topics in signal processing*, 1(4):586–597, 2007.

[9] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th international conference on international conference on machine learning*, pages 399–406, 2010.

[10] Rie Johnson and Tong Zhang. Accelerating stochastic gradient descent using predictive variance reduction. *Advances in neural information processing systems*, 26, 2013.

[11] Seung-Jean Kim, Kwangmoo Koh, Michael Lustig, Stephen Boyd, and Dimitry Gorinevsky. An interior-point method for large-scale l1-regularized least squares. *IEEE journal of selected topics in signal processing*, 1(4):606–617, 2007.

[12] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[13] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[14] Junfeng Yang and Yin Zhang. A fast iterative shrinkage-thresholding algorithm for compressive sensing. *IEEE Transactions on Signal Processing*, 58(5):2798–2805, 2010.