

# FINAL PROJECT BME 593

URI GOLDSZTEJN, MARIA MEDRANO, ESION LIN

## CONTENTS

1	Introduction	2
2	Methods	3
2.1	Stationary Methods: ART vs SART	3
2.2	Non-stationary methods: FISTA and NM	4
3	Results	9
3.1	Results of ART and SART	9
3.2	Optimization Results: FISTA	12
3.3	Optimization Results: New Method	14
4	Discussion	16
4.1	ART vs. SART vs. FISTA	16
4.2	FISTA vs. New Method	17
5	Conclusions	19
6	Appendix A: Derivation of New Method's Gradient Smooth Term	20
7	References	21

## LIST OF FIGURES

Figure 1	True Image	9
Figure 2	Convergence history of ART on 540 views	10
Figure 3	Convergence rate and scalability of ART	10
Figure 4	Convergence history of SART on 540 views	11
Figure 5	Convergence rate and scalability of SART	11
Figure 6	FISTA: L-Curve Criterion for Selecting Optimal Gamma	12
Figure 7	Reconstructed Image for Different Values of $\gamma_1$	13
Figure 8	FISTA: Convergence Rate Different Views	14
Figure 9	NM: Reconstructed Images Different Values of $\gamma_1$ and $\gamma_2$	15
Figure 10	New Method: Convergence Rate Different Views	15
Figure 11	Reconstructed Images	16
Figure 12	Reconstruction 90 Views FISTA vs. NM	18
Figure 13	Center Row Intensity FISTA vs. NM	19

## LIST OF TABLES

Table 1	ART vs SART: Iterations for an increasing number of views	11
Table 2	FISTA: Optimal $\gamma_1$ for Different Views	12
Table 3	FISTA: Iterations for Different Views	14
Table 4	New Method: Optimal $\gamma_1$ and $\gamma_2$ for Different Views	14
Table 5	New Method: Iterations for Different Views	15
Table 6	Mean Square Error: ART vs. SART vs. FISTA	16
Table 7	Mean Square Error: FISTA vs. New Method	18

## 1 INTRODUCTION

In Computed Tomography (CT) problems, we aim to restore the distribution of the attenuation coefficient in the target tissue. X-ray projections at uniformly spaced rotations are acquired in what is named the "forward" problem. The task is then to reconstruct the attenuation coefficient distribution that gave rise to the obtained projections, the "inverse problem". A classical approach is to use Filtered Back Projections (FBP) to solve the inverse problem. This approach, however elegant, is not robust to noise and gives unreliable solutions in the presence of noise or sparse acquisitions. As it is common with most inverse problems, the CT problem is ill-posed and additional conditions are required to find an optimal solution.

In this study, we explored two approaches to solving the CT inverse problem. Firstly, we explored stationary methods in which the problem is described as a linear system and the solution is truncated before convergence. This method presents a semiconvergent behaviour and the stopping criteria prevents overfitting to the noise. In particular, we explored the Algebraic Reconstruction Technique (ART), and an accelerated version, the Simultaneous Algebraic Reconstruction Technique (SART).

An alternative approach to solving inverse problem is to formulate them as optimization problems. This approach seeks to find a solution that explains the data observed (data fidelity term) and a given constrain, e.g. small norm or small gradients (regularization term). Both terms are balanced by regularization parameter which can then be tuned with different criteria. A common approach to solve the CT inverse problem is to use a regularization term that penalizes gradients. In particular, Total Variation (TV) has gained a lot of attention due to its edge preserving properties. TV regularized solutions give accurate results when sufficient samples are acquired and is currently one of the most popular approaches. However, for sparse acquisitions, angular aliasing presents upon reconstruction. Large spatial frequencies in the true image lead to radial "streaks" in the reconstructed image, when few projections are used. These streaks span radially from the high intensity components, and create angular discontinuities. Since TV addresses all the gradients in the image identically, it can not smooth this "streaks" without compromising the image quality.

Sparse acquisitions with reliable reconstructions are of particular interest. They allow for faster imaging with reduced radiation dose. These reconstructions can then, for example, be registered to higher quality reconstructions and used to detect anatomy changes, such as tumor growth or recession after therapy. Another promising use for sparse reconstructions is surgery guidance. If CT reconstructions could be processed in real time, they could be coupled with laparoscopic or robotic surgery, leading to a new gamut of possibilities.

In this study we propose a directional regularization term that penalizes angular gradient discontinuities and test its performance. This regularization term addresses the aliasing effect observed and allows for relaxation of the total variation penalization. The reconstruction quality is then enhanced by two means. The streaking artifact is reduced while the information containing edges are further preserved, thanks to a relaxed total variation term.

A key assumption of this method is that the high frequency objects tend to be centered in the field of view. This assumption holds in the CT inverse problem, since the target body is centered with respect to the rotating sensor array.

In this study, we used a synthetic thoracic CT and its simulated sinogram. The sinogram is corrupted with measurement noise as well as due to model inaccu-

racies. For the inverse problem formulation we used the linear imaging operator developed by Hansen et. al. We then proceed to implement four different methods to reconstruct the image using 90, 270, and 540 views, and compare the performance of each method from the least conditioned to the most ill-conditioned case. Finally, the performance of ART, SART, FISTA and the New Method are compared in terms of image quality, rate of convergence, and robustness.

## 2 METHODS

In order to solve this problem, we implemented a stationary and iterative method. In both instances the image was reconstructed using full tomographic views, half tomographic views, and one fourth tomographic views. The following are a description of each algorithm: their underlying mathematical principles, their implementation, and their parameter selection.

### 2.1 Stationary Methods: ART vs SART

**PROBLEM DESCRIPTION** Recall, the CT imaging system is modeled as a linear system:

$$d = Am$$

where  $d \in \mathbb{R}^m$  denotes the projection data,  $m \in \mathbb{R}^n$  attenuation coefficients over the region of interest, and  $A \in \mathbb{R}^{m \times n}$  the imaging operator. Expanding out the problem from the matrix formalism, we get  $m$  linear equations with  $n$  variables, with each equation spanning a hyperplane in  $\mathbb{R}^n$ . If there exists an exact solution to the problem, geometrically it should lie at the intersection of all hyperplanes.

Both stationary methods, ART and SART, attempt to locate the solution iteratively by projecting  $m$  onto the hyperplanes. However, the difference is that during each iteration, ART performs a round of projections where  $m$  is cast onto each hyperplane sequentially, while SART only takes one step by averaging all the steps required to project  $m$  onto all hyperplanes. Mathematically, the orthogonal distance from  $m$  to each hyperplane is  $A(i,:) \cdot m^{(k)} - d(i)$ , which should be subtracted during each projection operation. Finally, the norm difference between two iterations is computed to check for convergence. The iterations stop upon satisfying either one of the two criteria: (1) difference in  $\|m^{(k)}\|$  in the last 2 iterations is below an ad hoc tolerance, and (2) the number of iterations exceeds a preset maximum  $k_{max}$ .

#### PSEUDOCODE:

---

##### ART ( $A, d, k_{max}, tol$ )

Set  $m^{(k)} = 0, m^{(k)} \in \mathbb{R}^n$

Do while  $\delta(\|m\|) > tol^*$  and  $k \leq k_{max}^*$

For projection  $i \in [1, m]$  do

$$\delta(i) = d(i) - A(i,:) \cdot m^{(k)}$$

$$m^{(k)} = m^{(k)} + \delta(i) \cdot \frac{A(i,:)}{\|A(i,:)\|}$$

$$\delta(\|m\|) = \text{abs}(\|m^{(k)}\| - \|m^{(k-1)}\|)$$

$$\text{residual}^{(k)} = \|d - A \cdot m^{(k)}\|$$

$k = k + 1$

\*For the results presented below:  $k_{\max} = 100, \text{tol} = 50$

---

#### **SART ( $A, d, k_{\max}, \text{tol}$ )** [5] [6]

Set  $m_0 = 0, m_0 \in \mathbb{R}^n$

$\text{Sum}_{+,j} = \text{diag}(\|A(:,1)\|^{-1}, \dots, \|A(:,n)\|^{-1}) \in \mathbb{R}^{n \times n}$

$\text{Sum}_{i,+} = \text{diag}(\|A(1,:)\|^{-1}, \dots, \|A(m,:)\|^{-1}) \in \mathbb{R}^{m \times m}$

$A_{\text{norm}}^T = \text{Sum}_{+,j} \cdot A^T \cdot \text{Sum}_{i,+}$

Do while  $\delta(\|m\|) > \text{tol}^*$  and  $k \leq k_{\max}^*$

$\text{residual}^{(k-1)} = d - A \cdot m^{(k-1)}$

$m^{(k)} = m^{(k-1)} + A_{\text{norm}}^T \cdot \text{residual}^{(k-1)}$

$\delta(\|m\|) = \text{abs}(\|m^{(k)}\| - \|m^{(k-1)}\|)$

$k=k+1$

\* For the results presented below:  $k_{\max} = 300, \text{tol} = 6$

---

## 2.2 Non-stationary methods: FISTA and NM

**PROBLEM DESCRIPTION** As previously described, in order to reconstruct the X-ray CT image from an experimental data set, we must consider a discrete linear system of the form:

$$Ax + w = d$$

Where  $A \in \mathbb{R}^{m \times n}$  consists of the imaging operator (derived in the `paralleltomo.m` function),  $d \in \mathbb{R}^m$  is the acquired image,  $w$  is an unknown noise or perturbation vector, and  $x \in \mathbb{R}^n$  is the desired true image. However, unlike the previously described stationary method, in this instance the true image  $x$  will be found by solving a least square problem of the form.

$$\hat{x}_{LS} = \arg \min_x \|Ax - d\|^2$$

Since  $A$  can often be ill-conditioned, a regularization term is added to stabilize the solution. The Total Variation approach has gained much interest thanks to its edge preserving properties. This is particularly interesting in medical imaging, in which distinguishing key features is paramount.

The optimization problem then becomes:

$$\min_x \|\|Ax - d\|^2 + 2\gamma_1 \|x\|_{TV} \quad \text{Eq. (1)}$$

where  $A, x$ , and  $d$  are the same as defined above and  $\gamma_1 > 0$  is a penalization parameter providing a trade off between data fidelity and noise removal

Since it is composed of a smooth and a not smooth component, (1) cannot be solved using standard gradient descent. One way to tackle this is by replacing the non smooth version of TV for a smooth approximation of the form  $\sqrt{x + \xi}$ ; however, this method will lead to ill-conditioning as  $\xi \rightarrow 0$  and without proper step-size search methods, it can also be extremely slow. In order to avoid all these issues and to aim for high quality results, we decided to implement TV by using FISTA, a state of the art method for solving nonsmooth convex optimization problems. In this study, we also introduced an additional regularization term. This term aims

to reduce the streaking artifacts observed in sparse reconstructions by adding an additional cost to angular variations.

**ALGORITHM OVERVIEW** Beck and Teboulle [1], created the backdrop for FISTA by first presenting the iterative shrinkage algorithm (ISTA). The basic idea behind ISTA is to view each gradient iteration as a proximal regularization of a linearized function. In doing so, it is able to solve the nonsmooth  $l_1$  regularized problem through an iterative scheme which can be reduced to the solving a one dimensional minimization problem.

Considering this approximation, Beck and Teboulle proceeded to show an iterative scheme generalized to optimization problems of the form:

$$\min F(x) = f(x) + g(x) : x \in \mathbb{R}^n$$

where:

\* $g : \mathbb{R}^n \rightarrow \mathbb{R}$  is a continuous convex function which can possibly be nonsmooth.

\* $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a smooth convex function of the type  $C^1$ .

\*and the problem is solvable. In other words  $X_* := \operatorname{argmin} F \neq \emptyset$  and for  $x^* \in X_*$  we set  $F_* := F(x^*)$ .

Based on ISTA's approximation scheme, for any  $L > 0$ , Beck and Teboulle then consider the quadratic approximation at a given point  $y$  as:

$$Q_L(x, y) := f(y) + \langle x - y, \nabla f(y) \rangle + \frac{L}{2} \|x - y\|^2 + g(x)$$

Which admits a unique minimizer:

$$p_L(y) := \arg \min Q_L(x, y) : x \in \mathbb{R}^n$$

That can also be rewritten as:

$$p_L(y) = \arg \min_x \left[ g(x) + \frac{L}{2} \left\| x - \left( y - \frac{1}{L} \nabla f(y) \right) \right\|^2 \right] \quad \text{Eq. (2)}$$

Thus, based on ISTA's model, the basic step for the generalized model can then be simply written as:

$$x_k = p_L(x_{k-1})$$

where  $L$  plays the role of a stepsize and where, in the case of our implementation, we consider  $g(x)$  to be Total Variation regularization and  $f(x)$  to be the data fidelity term  $\frac{1}{2} \|Ax - d\|^2$ .

Beck and Teboulle then ingeniously combine the proximal regularization from the generalized model based on ISTA with the accelerated gradient scheme derived by Nesterov in [3] to then create an algorithm capable of solving non smooth convex problems with an improved complexity of  $O(1/k^2)$ , FISTA.

In implementing FISTA; however, we still face the challenge of finding a method to determine the proximal for the Total Variation regularizer. Therefore, we referred to a second paper by Beck and Teboulle [2] where problems of the form:

$$\min_{x \in C} \frac{1}{2} \|x - b\|_F^2 + 2\gamma_1 \operatorname{TV}(x)$$

such as Eq. (2) can be solved by considering a dual approach and then by using a gradient-based algorithm to solve it. In [2], the Fast Gradient Proximal algorithm (FGP) solves for the proximal of TV by using gradient descent combined with Nesterov's accelerated gradient to find a pair  $(p, q) \in P$  that will be an optimal solution

to the dual problem.

$$\min_{(p,q) \in P} \left[ h(p, q) = -\|H_C(b - \gamma_1 L(p, q))\|_F^2 + \|b - \gamma_1 L(p, q)\|_F^2 \right]$$

Where

$$H_C(x) = x - P_C(x) \text{ for every } x \in \mathbb{R}^{m \times n}$$

and for which then the optimal solution for the problem in Eq. (2) is of the form:

$$x = P_C(b - \gamma_1 L(p, q))$$

Therefore, based on the derivations in [1] and [2] we decided to

- (1) Use the proximal regularization update on Eq. (2) with  $g(x) = \gamma_1 (TV)$  to update  $x_k$ .
- (2) Determine the proximal  $p_L(x_{k-1})$  by using a dual approach and an accelerated gradient descent (FGP) as described on [2].
- (3) and use Nesterov's accelerated gradient method to have an improved complexity result of  $O\left(\frac{1}{k^2}\right)$  for each FISTA iteration

In doing so we were then able to implement in a reasonable amount of iterations an algorithm that would remove extraneous noise while still preserving the edges when reconstructing a corrupted CT image  $x$ .

**NEW METHOD** In general, FISTA is considered as a state of the art method that provides superior results in a reasonable amount of time. However, as the sampling views is reduced aliasing becomes apparent and none of the aforementioned methods, not even FISTA, is able to remove the streaks without blurring the target image. In particular, increasing the regularization parameter  $\gamma_1$  leads to a complete loss of detail in the reconstructed image before achieving streak removal. To address this issue, we propose an additional regularization term that targets the aliasing noise. This terms penalizes directional derivatives in the angular directions.

The additional term is given by  $\int_{\Omega} \nabla m(H(x) \cdot D(x) \cdot \nabla m) \, dx$  Where  $m$  is the current image iterate,  $\Omega$  is the image region,  $D(x) = d(x) \otimes (x)$ , where  $d(x)$  is a unit vector in the angular direction, i.e.  $d(x) = [-\sin(\alpha), \cos(\alpha)]$ . Lastly,  $H$  is a linear radial filter that compensates the concentration of smoothing directions towards the center of the image. It is defined as.

$$H(x) = \min \left( 1, \frac{\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}}{n} \right)$$

where  $(x_0, y_0)$  is the center of the image and  $x = (x_1, y_1)$ .

Since this additional term is quadratic, it is differentiable. Its gradient can then be added to the gradient of the data fidelity term and plugged into the FISTA routine for optimization.

By adding our new regularization term to the problem originally solved by FISTA our new optimization problem becomes:

$$\min_{x \in C} \frac{1}{2} \|Ax - d\|^2 + \gamma_2 \cdot \int_{\Omega} \nabla m(H(x) \cdot D(x) \cdot \nabla m) \, dx + \gamma_1 TV(x) \quad \text{Eq. 3}$$

Where the gradient of the smooth portion of our objective function is then given by:

$$\nabla f = A'(Ax - d) - \nabla(H(x) \cdot D(x) \cdot \nabla m)$$

This gradient can be found using matrix differentiation and calculus of variations. The derivation is shown in the Appendix A.

**PSEUDOCODE** Based on the derivations from Beck and Teboulle in [1] and [2], we proceeded to implement FISTA and FISTA with the the proposed additional regularization term to reconstruct a set of CT images by using 90, 270 and 540 tomographic views. The pseudocode for our implementation is as follow.

---

#### Function for Calculating $\nabla f(m)$

\*This function is only used when implementing our new method to calculate b in FISTA's algorithm

Set

$m_k \in \mathbb{R}^{n \times n}$  - Current image iterate  
 $d\theta \in \mathbb{R}$  -Angular spacing between views in radians:  
 $m_x$  -Column-Wise Difference of  $m_k$   
 $m_y$  -Row-Wise difference of  $m_k$   
 $x_0$  -Round( $n/2$ )  
 $y_0$  -Round ( $n/2$ )

For i in 1 to n do

For j in 1 to n do

$$\alpha = \arctan(y - y_0)/(x - x_0)$$

if  $y - y_0 < 0$

$$\alpha = \alpha + \pi \hat{\alpha}^*$$

$$d(i, j) = [-\sin(\hat{\alpha}), \cos(\hat{\alpha})]$$

$$D(i, j) = d(i, j)' \cdot d(i, j)$$

$$H(i, j) = \min(1, \sqrt{(j - x_0)^2 + (i - y_0)^2}/n)$$

$$H\nabla m(i, j) = H(i, j) \cdot D(i, j) \cdot [m_x, m_y]^T$$

$F_x^{**}$

$F_y^{***}$

$$\nabla f = A'(Ax - d) - \gamma_2 \cdot (F_x + F_y)$$

\* closest multiple of  $d\theta$  to  $\alpha$

\*\* reverse column-wise difference of the first elements of  $H \nabla m$

\*\*\*reverse row-wise difference of the second elements of  $H \nabla m$

---



---

#### FISTA Algorithm

Set  $y_1 = x_0, t_1 = 1, x_0 \in \mathbb{R}^n, L = 1$

For  $\|x_k - x_{k-1}\| \leq \xi$  do

Set  $b = y_k - A'(Ay_k - d)$  \*

Take  $(r_1, s_1) = (p_0, q_0) = (0_{(m-1) \times n}, 0_{m \times (n-1)})$  and  $t_1 = 1$

For Iteration= 1 to Max Iterations N do \*\*

$$(p_k, q_k) = P_p[(r_k, s_k) + \frac{1}{8\lambda} L^T (P_c(b - \lambda L(r_k, s_k)))]$$

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

$$\begin{aligned}
(r_{k+1}, s_{k+1}) &= (p_k, q_k) + \left( \frac{t_k - 1}{t_{k+1}} \right) (p_k - p_{k-1}, q_k - q_{k-1}) \\
x_k &= P_C[b - \lambda L(p_N, q_N)] \\
t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2} \\
y_{k+1} &= x_k + \left( \frac{t_k - 1}{t_{k+1}} \right) (x_k - x_{k-1})
\end{aligned}$$

\* For FISTA:  $b = y_k - A'(Ay_k - d)$ .

For New Method:  $b = A'(Ax - d) - \gamma_2 \cdot (F_x + F_y)$

\* Refer to **Function for Calculating**  $\nabla f(m)$  for details

\*\*For this implementation we used [4] to perform the inner loop calculations

Where the inner loop in **FISTA Algorithm** estimates the proximal of TV by using the FGP algorithm in [2] where

\*  $L : \mathbb{R}^{(m-1) \times n} \times \mathbb{R}^{m \times (n-1)} \rightarrow \mathbb{R}^{m \times n}$  is the linear operator defined by

$$L(p, q)_{i,j} = p_{i,j} + q_{i,j} - p_{i-1,j} - q_{i,j-1} \quad i = 1, \dots, m, j = 1, \dots, n$$

where  $p_{0,j} = p_{m,j} = q_{i,0} = q_{i,n} = 0$  for every  $i = 1, \dots, m$  and  $j = 1, \dots, n$ .

\*  $L^T : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{(m-1) \times n} \times \mathbb{R}^{m \times (n-1)}$  is the adjoint to  $L$  and is defined by  $L^T(x) = (p, q)$ .

\*  $p \in \mathbb{R}^{(m-1) \times n}$  and  $q \in \mathbb{R}^{m \times (n-1)}$  are defined by

$$p_{i,j} = x_{i,j} - x_{i+1,j}, \quad i=1, \dots, m-1, j=1, \dots, n$$

$$q_{i,j} = x_{i,j} - x_{i,j+1}, \quad i=1, \dots, m, j=1, \dots, n-1$$

\*  $P_C$  is the orthogonal projection operator onto the set  $C$ .

\*  $P_p$  is a projection to the set  $P$ . For a pair  $(p, q)$ ,  $P_p(p, q) = (r, s)$

\* The discrete TV function used is the isotropic TV which is.

$$\begin{aligned}
x \in \mathbb{R}^{m \times n}, \quad TV_I(x) &= \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sqrt{(x_{i,j} - x_{i+1,j})^2 + (x_{i,j} - x_{i,j+1})^2} \\
&+ \sum_{i=1}^{m-1} |x_{i,n} - x_{i+1,n}| + \sum_{j=1}^{n-1} |x_{m,j} - x_{m,j+1}|
\end{aligned}$$

And for which,  $r \in \mathbb{R}^{(m-1) \times n}$ , and  $s \in \mathbb{R}^{m \times (n-1)}$  are defined as.

$$\begin{aligned}
r_{i,j} &= \begin{cases} \frac{p_{i,j}}{\max[1, \sqrt{p_{i,j}^2 + q_{i,j}^2}]}, & i = 1, \dots, m-1, j = 1, \dots, n-1 \\ \frac{p_{i,n}}{\max[1, |p_{i,n}|]}, & i = 1, \dots, m-1 \end{cases} \\
s_{i,j} &= \begin{cases} \frac{q_{i,j}}{\max[1, \sqrt{p_{i,j}^2 + q_{i,j}^2}]}, & i = 1, \dots, m-1, j = 1, \dots, n-1 \\ \frac{q_{m,j}}{\max[1, |q_{m,j}|]}, & j = 1, \dots, n-1 \end{cases}
\end{aligned}$$

**PARAMETER SELECTION** In this particular implementation,

(1) The optimal  $\gamma_1$  for FISTA and  $\gamma_1, \gamma_2$  for our New Method for each one of the three different views was selected through an L-curve criterion. The optimal  $\gamma_1$  for 90, 270 and 540 views are presented on Table 2 and the L-curve criterion for each case are shown on Figure 6. The optimal  $\gamma_1, \gamma_2$  for the new method for three different views will be shown in Table 4.

(2) The value of  $L$  used in **Eq. (2)** for each  $x_k$  update in FISTA was fixed to 1 to



make a direct implementation of FGP in [2] for solving a problem of the form  $\min_{x \in C} \frac{1}{2} \|x - b\|_F^2 + 2\lambda \text{TV}(x)$ . Although we sacrificed an even faster convergence to simplify our implementation, FISTA with a fixed  $L$  value was still able to converge to a solution in a reasonable number of iterations as will be seen in the Results section.

(3) The stopping condition for both methods was determined by the norm difference between  $x_k$  and  $x_{k-1}$ . The algorithm converged when the norm difference between both was less than a threshold  $\xi$ . In this case  $\xi$  was selected to be 5, in order to allow the algorithm to finish in a reasonable amount of time and still get comparable results between different methods.

### 3 RESULTS

Now, we shall describe the optimal parameter and convergence rate when reconstructing the X-ray tomographic images using 90, 270, and 540 tomographic views. The following figure is the original X-ray CT image that will be reconstructed using the four methods and will serve as a ground truth reference for all subsequent reconstructions shown in this report.

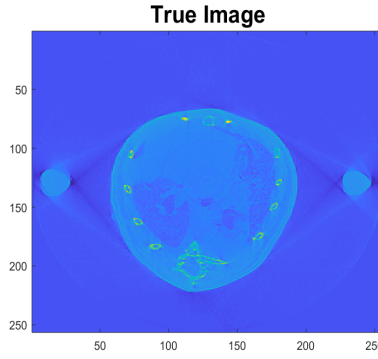


Figure 1

#### 3.1 Results of ART and SART

##### 3.1.1 ART

The residual history on Figure 2 suggests that ART converged in around 30 iterations. As shown in Figure 2,  $m^{(k)}$  oscillated between iterations before stabilizing to a relatively constant magnitude at 50 iterations, confirming the stopping criterion. Further, convergence rate as a function of the number of views is illustrated in Figure 3. A rapid increase in the number of iterations was not observed as the number of views increased, suggesting that the algorithm is relatively scalable. However, the plot only elucidates the iterations in the outer while loop—referring back to the ART pseudocode—while the computation of the inner for loop expanded directly with the number of views. In this implementation no parameter selection was required, and the algorithm stopped when the norm difference between  $m^{k-1}$  and  $m^k$  was lower than a given threshold ( $\text{tol} = 50$ ) or when the number of iterations reached a maximum ( $k_{\max} = 100$ ).

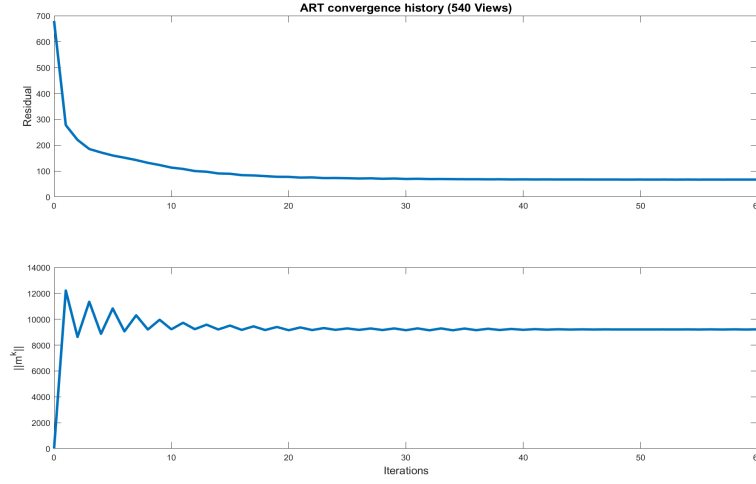


Figure 2: Convergence history of ART on 540 views

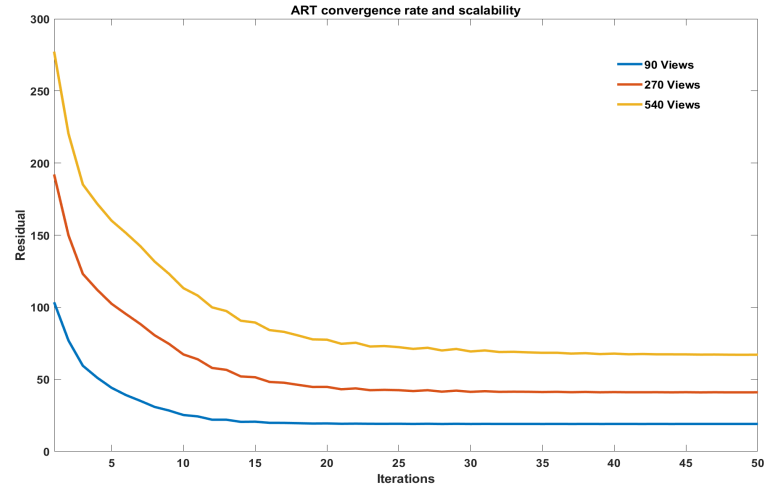


Figure 3: Convergence rate and scalability of ART

### 3.1.2 SART

As shown in Figure 4, the convergence of  $m^{(k)}$  towards a relatively constant value in SART was smoother than ART, but took significantly more iterations (150 vs 50). As described in the methods sections, SART also stopped when the norm difference between  $m^{k-1}$  and  $m^k$  was lower than a given threshold ( $\text{tol} = 6$ ) or when the number of iterations reached a maximum ( $k_{\max} = 300$ ). The convergence history for different number of views is illustrated in Figure 5.

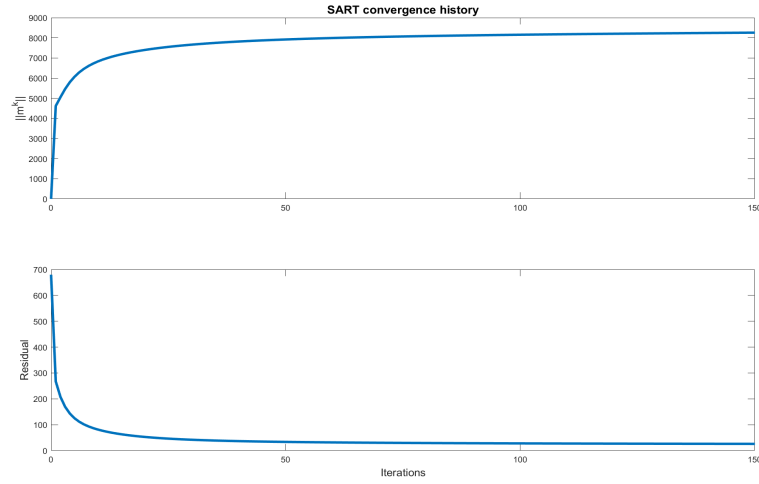


Figure 4: Convergence history of SART on 540 views

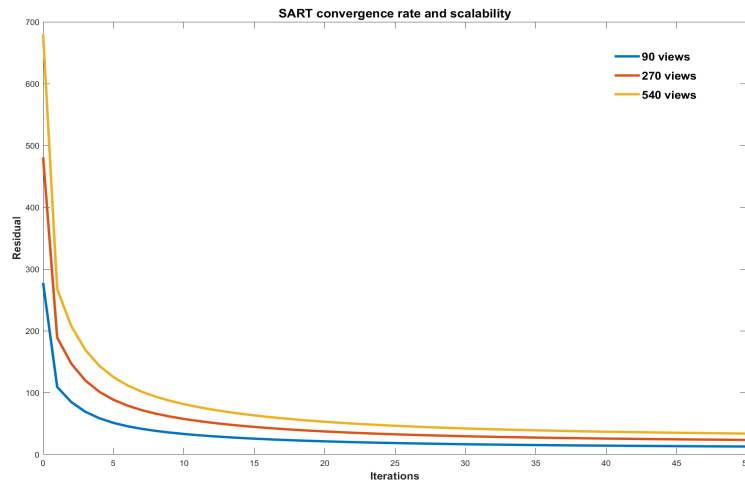


Figure 5: Convergence rate and scalability of SART

Table 1 summarizes the number of iterations for convergence with respect to an increased number of views in ART's and SART's implementation. As previously mentioned ART was able to converge with much less iterations than SART. A comparison in the quality of both reconstructed images will be shown later in the Discussion section.

Table 1: ART vs SART: Iterations for an increasing number of views

Views	ART	SART
90 Views	39	146
270 Views	55	162
540 Views	74	189

### 3.2 Optimization Results: FISTA

#### 3.2.1 L-Curve Criterion for Different Views

For determining an optimal  $\gamma_1$  parameter for FISTA, we only required to record the final data misfit and regularization when using 9 different values of  $\gamma_1$  and then apply the L-curve criterion to select the optimal parameter value. Figure 5 and Table 2 display the L-curves used for making this parameter selection and summarize the final optimal gamma for each set of views.

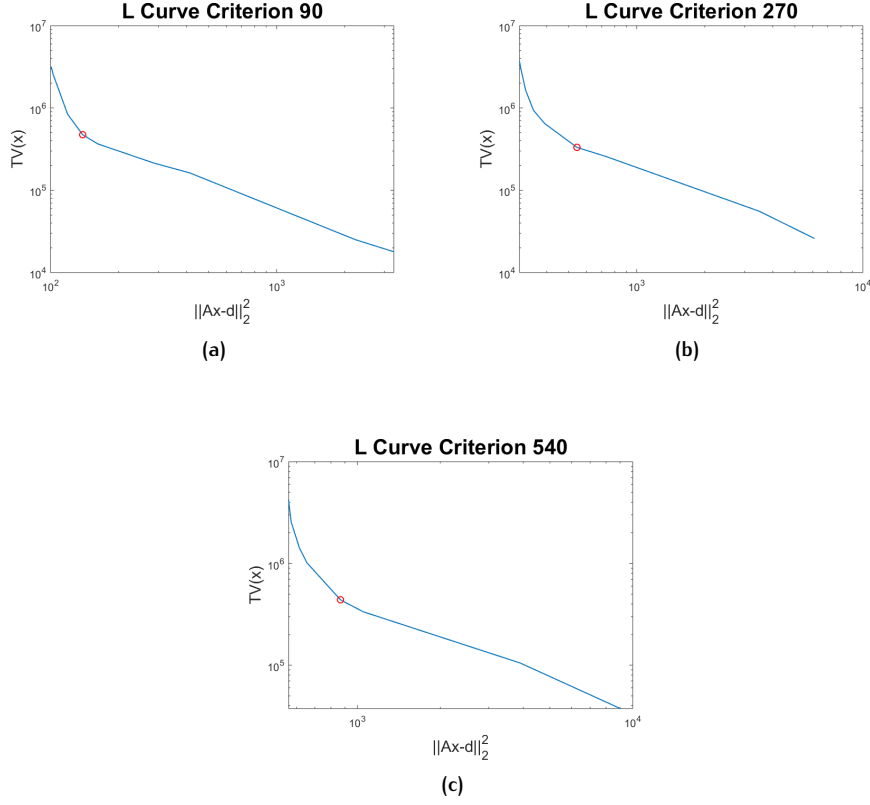


Figure 6: FISTA: L-Curve Criterion for Selecting Optimal Gamma

Table 2: FISTA: Optimal  $\gamma_1$  for Different Views

Views	Optimal $\gamma_1$
90 Views	0.0001
270 Views	0.0001
540 Views	0.0001

From Table 2, we can see that the best value of  $\gamma_1$  for all three views was 0.0001. This was further verified by observing the reconstructed images for a higher and a lower value of  $\gamma_1$ . The true image can be referenced in Figure 1.

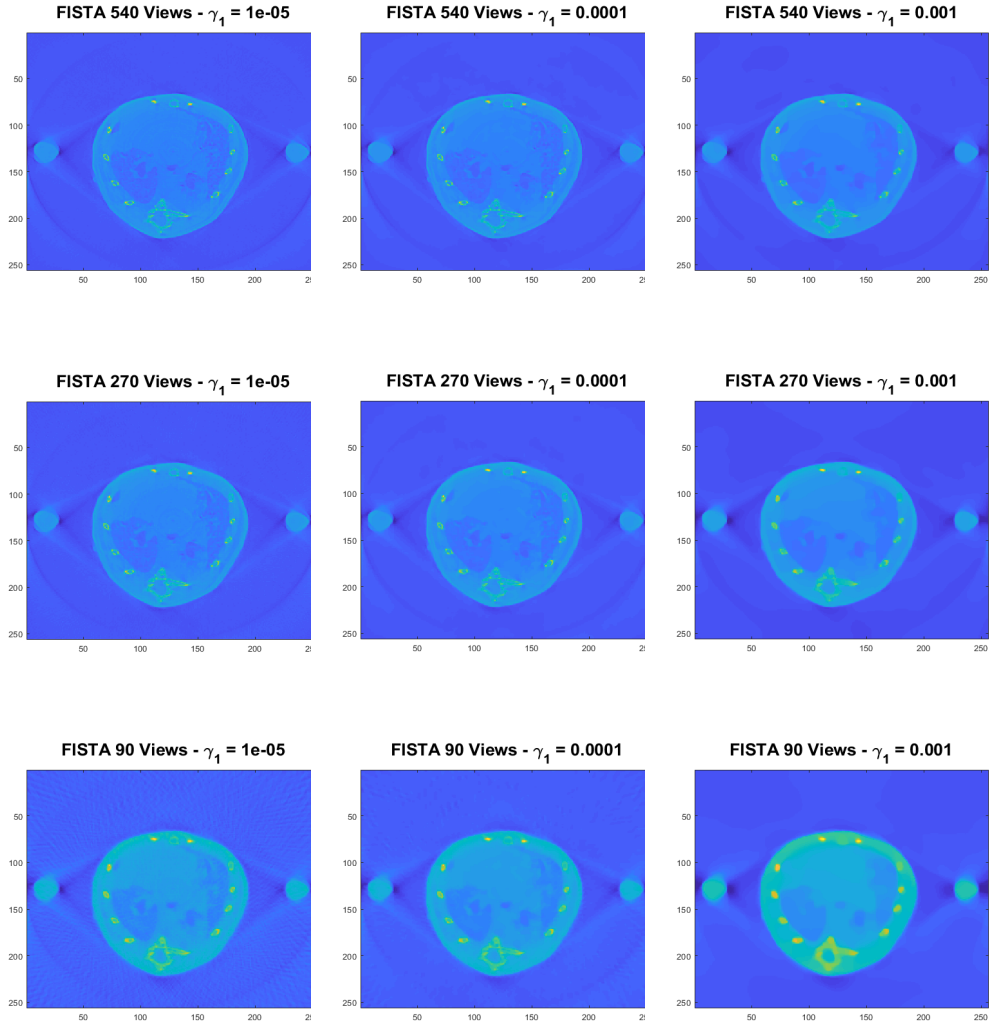


Figure 7: Reconstructed Image for Different Values of  $\gamma_1$

As it can be observed in Figure 7, a higher value of  $\gamma_1$  led to more smoothing and worst edge preservation—especially in the case of 90 views—and a lower value of  $\gamma_2$  led to more details and noise. Our L-Curve criterion, though, provided an optimal value of  $\gamma_1$  that allowed enough smoothing without losing the central sharp edges in the figure.

### 3.2.2 Convergence Rate

Using the values of  $\gamma_1$  and  $\gamma_2$  estimated on the previous subsection, we then proceeded to implement FISTA to reconstruct a series of CT images with different number of tomographic views. From Figure 8, we notice that the rate of convergence of our implementation for 3 different views was superlinear or approximately quadratic; thus, verifying that the rate of convergence of our algorithm was superior to the linear rate of convergence of a simple gradient descent.

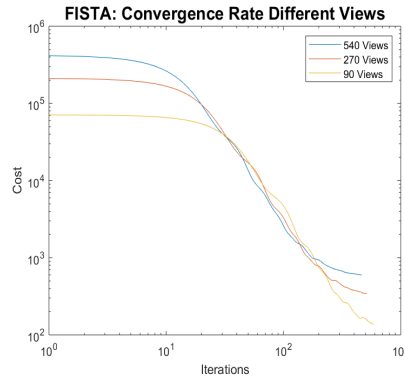


Figure 8

Table 3 summarizes the number of iterations the algorithm required to converge for an increasing number of tomographic views.

Table 3: FISTA: Iterations for Different Views

Views	Iterations
90 Views	591
270 Views	514
540 Views	466

From Table 3, it can be observed that for an increasing number of tomographic views, FISTA required a lower number of iterations. This is a reasonable behavior; since, as the number of tomographic views decreases, the problem becomes more ill-conditioned, and FISTA requires more iterations to converge to a solution.

### 3.3 Optimization Results: New Method

#### 3.3.1 L-Curve Criterion for Different Views

The following table summarizes the ideal  $\gamma_1$  and  $\gamma_2$  values used for reconstructing different tomographic views. In this case, unlike FISTA, we required a total of  $3 \times 18$  L-Curves. For each L-Curve we ran our new method using one fixed  $\gamma$ —either  $\gamma_1$  or  $\gamma_2$ —value and used 9 different values for the unfixed  $\gamma$ . Thus, we ran the algorithm for 81 different  $\gamma$  combinations for each set of views. For the sake of conciseness we don't show all the L-Curves; instead, Table 4 summarizes the parameters that were determined by the L-Curve Criterion and later verified through the reconstructed images.

Table 4: New Method: Optimal  $\gamma_1$  and  $\gamma_2$  for Different Views

Views	$\gamma_1$	$\gamma_2$
90 Views	0.0001	0.0001
270 Views	0.0001	0.0001
540 Views	0.0001	0.0001

As shown in Table 4, The optimal values for  $\gamma$  were the same for each set of views. Additionally, just as it was seen in FISTA, a higher  $\gamma$ —either  $\gamma_1$  or  $\gamma_2$ —led to more smoothing and a lower  $\gamma$  led to more noise. This was more apparent in the case of 90 views as can be observed in Figure 9.

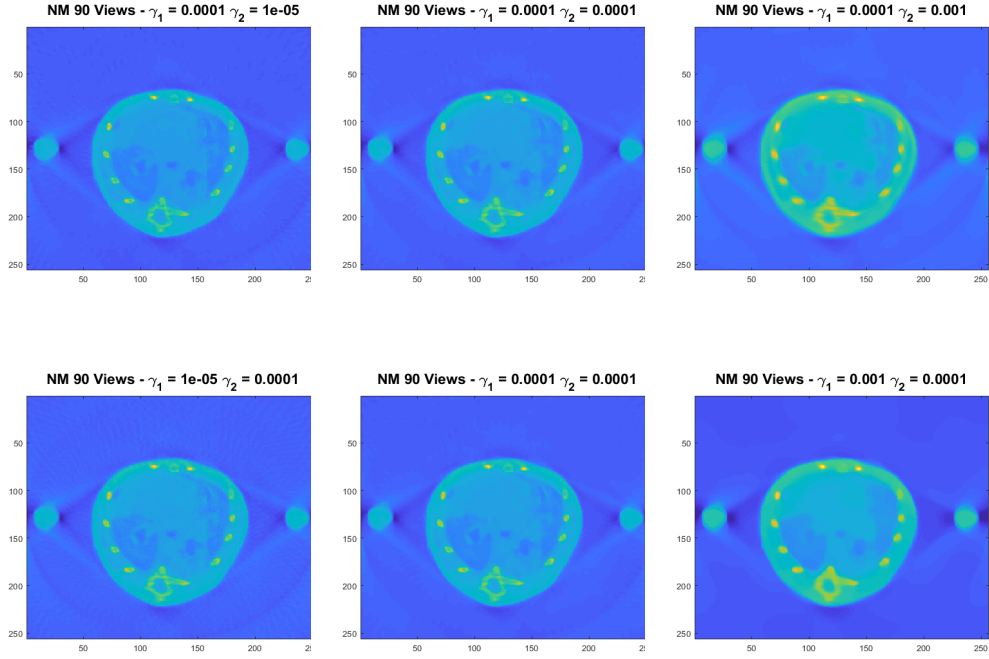


Figure 9: NM: Reconstructed Images Different Values for  $\gamma_1$  and  $\gamma_2$

### 3.3.2 Convergence Rate

Using the values of  $\gamma_1$  and  $\gamma_2$  estimated on the previous subsection, we then proceeded to implement our new method to reconstruct a series of CT images with different number of tomographic views. As it can be observed in Figure 10, the rate of convergence for our New Method was also superlinear or approximately quadratic.

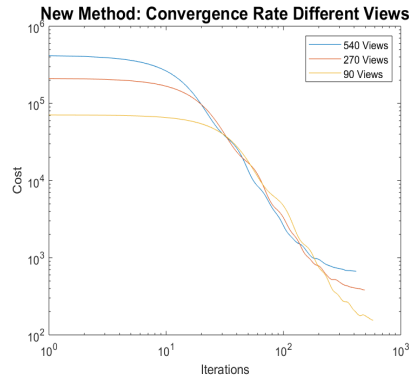


Figure 10

Table 5: New Method: Iterations for Different Views

Views	Iterations
90 Views	582
270 Views	493
540 Views	418

It can also be noticed in Table 5 that, just as in FISTA, the number of iterations decreased as the number of views increased. Again, this can be attributed to the ill-conditioning that becomes more prevalent as the number of views decreases.

## 4 DISCUSSION

### 4.1 ART vs. SART vs. FISTA

The following is a final overview of the performance of each method when using different tomographic views for reconstruction. The main metrics for testing performance in this case were mean square error between the reconstructed and the reference image and the visual appearance of the final reconstruction. The following figure show the resulting reconstructed images for different numbers of views (each row) and different methods (each column), and the subsequent table then summarizes the MSE that resulted for each set of methods and views. The true image in Figure 1 serves as a reference.

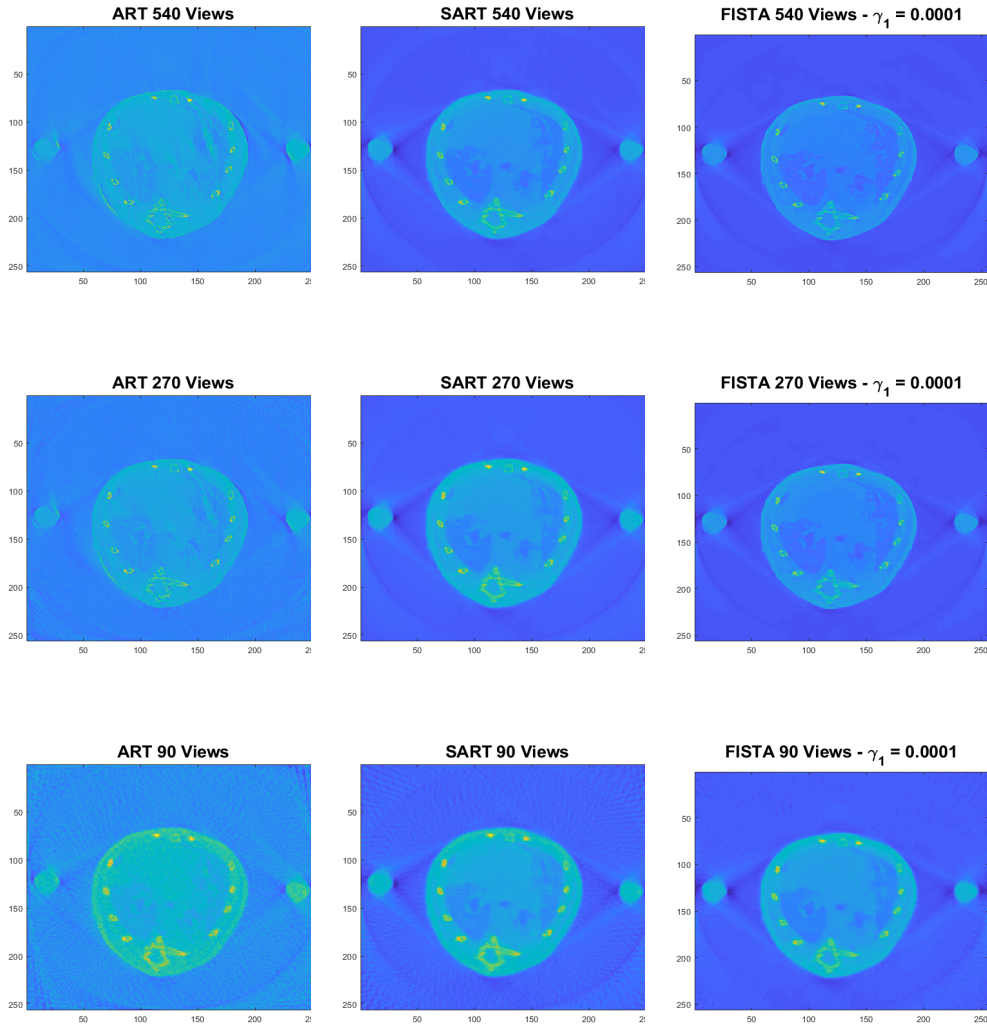


Figure 11: Reconstructed Images: 540 Views

Table 6: Mean Square Error: ART vs. SART vs. FISTA

Views	ART	SART	FISTA
90 Views	$4.55 \times 10^3$	$3.11 \times 10^3$	$2.05 \times 10^3$
270 Views	$4.29 \times 10^3$	$1.64 \times 10^3$	$1.48 \times 10^3$
540 Views	$3.31 \times 10^3$	$1.96 \times 10^3$	$1.29 \times 10^3$



On Figure 11 and Table 6, it can be seen that the difference in performance between ART, SART, and FISTA became more apparent as the number of tomographic views decreased. In the case of 540 tomographic views, the performance of all four methods was quite close to each other. The three of them preserved well the edges and the intensity of the reconstructed image was close to the desired one. The only major difference was seen in ART which failed to preserve most of the intensity and contrast from the original image. SART, on the other hand, was far superior than ART in reducing artifact and showing contrast for all three views; clearly this supports the observation made on Homework 2 in which we discovered SART was more robust than ART to noise corrupted data.

Once we reduced the number of tomographic views to 270, ART's performance deteriorated even further. SART and FISTA, on the other hand, proved to be robust to the discontinuities in the data caused by a decreasing number of views and were able to reconstruct an image relatively close to the desired one. As seen on Figure 11, SART and FISTA remained quite unaffected by the number of tomographic views while ART started showing a few uneven intensity changes throughout the image.

When we reduced the number of views for reconstruction to only 90, the inverse problem became even more ill-conditioned and the difference in performance between Stationary and Non-Stationary Methods was more apparent. As it can be seen on Figure 11, it is quite clear that FISTA with an appropriate gamma value far outperformed SART and ART in reproducing the intensity and edges of the original image. By using the tuning parameter  $\gamma_1$  and the edge preserving regularizer TV, FISTA was able to reconstruct an image while maintaining a trade-off between data fidelity and noise removal.

It is also worth pointing out that for each method there was a far superior performance as the number of tomographic views increased. A higher number of tomographic views led to a lower MSE for each one of the methods. This behavior can also be seen across each one of the rows on Table 6. Going from ART to SART and then to FISTA we notice a substantial decrease in MSE. In Table 6 we can further notice that the image with the lowest MSE was reconstructed by FISTA with 540 views, and the image with the highest MSE was reconstructed by ART with 90 views.

#### 4.2 FISTA vs. New Method

For this project we also explored a new regularization scheme that would aim to remove streak artifacts caused by missing data while still preserving the edges. As was observed on Figure 11, the biggest challenge for all four methods was in reconstructing the image in the most ill-conditioned scenario (90 views) and FISTA in this case had the best performance. Therefore it is worth placing emphasis in the comparison between our proposed algorithm and FISTA when using 90 views. Figure 12 and Table 8, demonstrate the performance of both methods based on image quality and MSE.

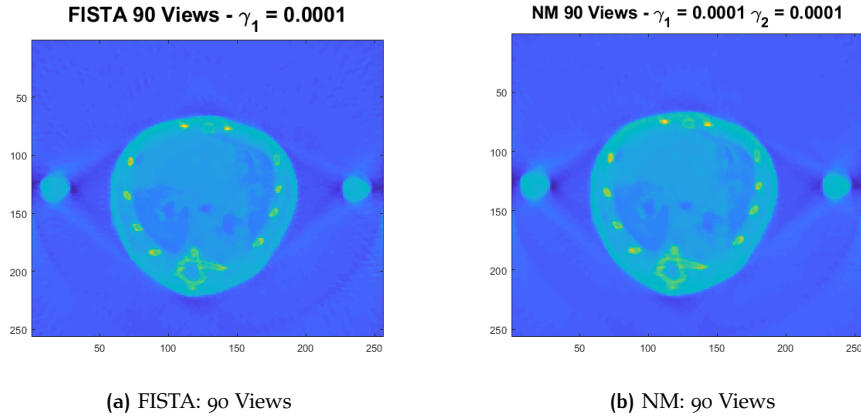


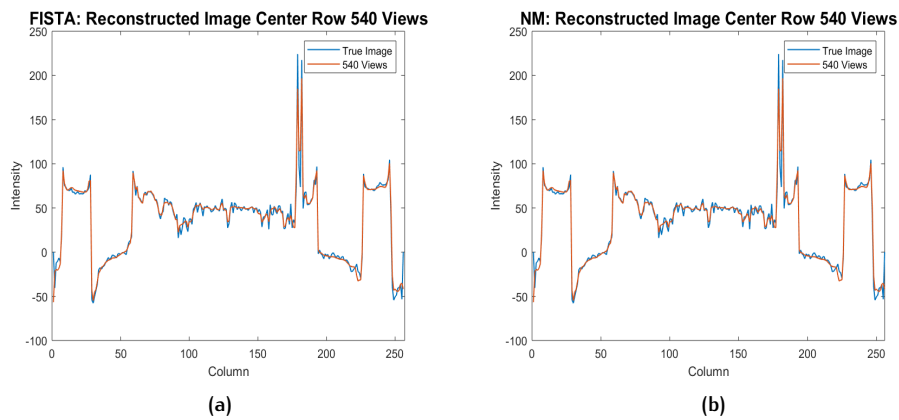
Figure 12: Reconstructed Images: 90 Views FISTA vs. NM

Table 7: Mean Square Error: FISTA vs. New Method

Views	FISTA	New Method
90 Views	$2.05 \times 10^3$	$2.13 \times 10^3$
270 Views	$1.48 \times 10^3$	$1.66 \times 10^3$
540 Views	$1.29 \times 10^3$	$1.47 \times 10^3$

By observing Figures (a) and (b) and comparing MSE in Table 8, it might seem there is not much difference between FISTA and our New Method; in fact the lower MSE might actually imply that FISTA is better. However, when observing the outer corners of both images, we notice that NM with its ideal values for  $\gamma_1$  and  $\gamma_2$  was able to suppress more the slight streak artifacts which FISTA failed to remove. It can also be seen that this improved streak suppression did not lead to further smoothing of the central image features. Overall, despite its slightly higher MSE value, the new method was able to more strategically suppress the noise without affecting the regions of interest from the original image.

In addition to analyzing MSE and reconstructed image quality, we also compared FISTA and the proposed method by looking at the resulting intensities in the image's central row.



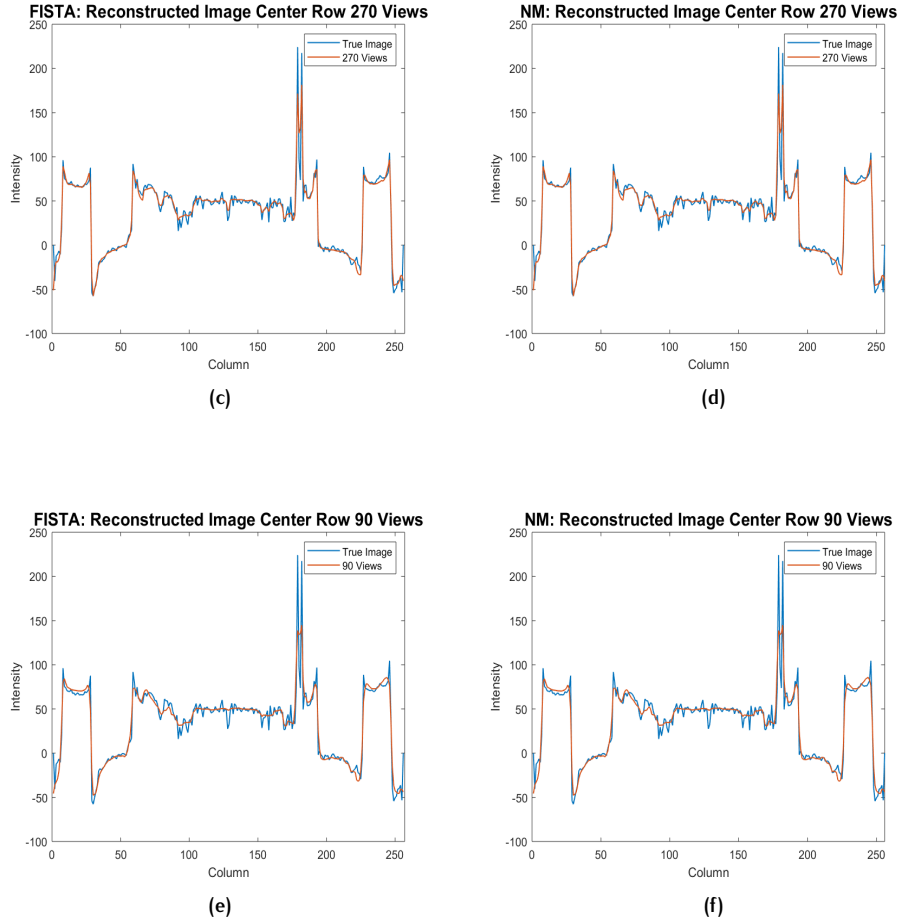


Figure 13: Center Row Intensity FISTA vs. NM

As shown in Figure 13, the contour of the center row intensity of both methods was not different for 540, 270, and 90 set of views. Images (a) and (b) — corresponding to 540 tomographic view reconstruction — most closely followed the spikes as well as the general contour of the true image at its center row. Later on, though, as the number of views decreased, the harder it became for the reconstructed image to fully recreate the original. In images (e) and (f) we notice that the intensity of the central row of the image reconstructed using 90 views followed the general contour of the original image, but without most of its spikes. Overall, this shows that for both methods, the lower the number of views, the more smoothing and the lower the edge preservation. Additionally, it also demonstrates that our New Method was able to attain the same amount of edge preservation as FISTA in the image's central edges.

## 5 CONCLUSIONS

In the Results and Discussion section, we observed that the stationary methods provided reliable solutions. However, their convergence was slow and the regularization of the solution could not be easily controlled. FISTA provided a faster convergence rate and yielded better results in terms of convergence speed and MSE, furthermore, in Figure 13, we observed that the solution obtained with FISTA accurately captured the edges of the true image.

However, although these methods yielded reliable solutions for 540 views, none of them reduced the aliasing artifacts observed in the sparse reconstructions. In particular, increasing the TV regularization parameter led to a complete loss of detail in the target image before removing these artifacts. Our proposed method, on the other hand, was able to reduce the aliasing artifacts without compromising the details in the target image. These can be observed in the reconstructed images as well as in the plots for the middle row intensities. Thus, our method, unlike the other three, was able to effectively remove the streak artifacts while not compromising the image's details.

It is left for future studies to investigate a non-smooth version of the proposed regularization term, i.e.  $g(x) = \gamma_2 \cdot \int_{\Omega} \sqrt{\nabla m(H(x) \cdot D(x) \cdot \nabla m)} dx$  and to further analyze its performance to highly noisy and ill-conditioned cases.

## 6 APPENDIX A: DERIVATION OF NEW METHOD'S GRADIENT SMOOTH TERM

The smooth term in the final objective function in [Eq. 3](#) is

$$f(x) = \frac{1}{2} \|Ax - d\|^2 + \gamma_2 \cdot \int_{\Omega} \nabla m(H(x) \cdot D(x) \cdot \nabla m) dx + \gamma_1 TV(x)$$

From previous derivations done in class we know

$$\frac{d}{dx} \frac{1}{2} \|Ax - d\|^2 = A'(Ax - d) \quad (*)$$

Therefore we focus on determining:

$$\frac{d}{dx} \nabla m(H(x) \cdot D(x) \cdot \nabla m)$$

It then follows:

$$\begin{aligned} & \frac{d}{dx} \nabla m(H(x) \cdot D(x) \cdot \nabla m) \\ &= \lim_{\xi \rightarrow 0} \frac{f(m + \xi \tilde{m}) - f(m)}{\xi} = \lim_{\xi \rightarrow 0} \frac{\nabla(m + \xi \tilde{m}) [H \cdot \nabla(m + \xi \tilde{m})] - \nabla m [H \cdot \nabla m]}{\xi} \\ &= \lim_{\xi \rightarrow 0} \frac{\nabla m \cdot H \cdot D \cdot \nabla m + 2\xi \nabla \tilde{m} \cdot H \cdot D \cdot \nabla m + \nabla(\xi \tilde{m}) \cdot H \cdot D \cdot \nabla(\xi \tilde{m}) - \nabla m \cdot H \cdot D \cdot \nabla m}{\xi} \\ &= \lim_{\xi \rightarrow 0} 2\nabla \tilde{m} \cdot \nabla m + \xi \nabla \tilde{m} = 2\nabla \tilde{m} \cdot (H \cdot D \cdot \nabla m) \end{aligned}$$

Using the Dirichlet boundary condition and Green's Identity it follows that:

$$\int_{\Omega} 2\nabla \tilde{m} \cdot (H \cdot D \cdot \nabla m) dx = - \int_{\Omega} 2\nabla \cdot (\nabla(H \cdot D \cdot m)) dx + \int_{2\Omega} \tilde{m} \cdot \nabla m dx$$

And  $\tilde{m} = 0$  on  $2\Omega$

$$\Rightarrow \nabla \tilde{m} (H \cdot D \cdot \nabla m) = \nabla \cdot (\nabla(H \cdot D \cdot m)) = -\nabla(\nabla H \cdot \nabla m) = -\nabla(H \cdot D \cdot \nabla m) \quad (**)$$

Considering results (\*) and (\*\*) it then follows:

$$\nabla f(x) = A'(Ax - d) - \gamma_2 \nabla(H \cdot D \cdot \nabla m)$$

## 7 REFERENCES

- [1] Amir Beck, Marc Teboulle. *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems*. SIAM J. Imaging Sciences, Vol. 2, No.2, pp. 183-202.
- [2] Amir Beck, Marc Teboulle. *Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems*. IEEE Transactions on Image Processing, Vol. 18, No. 11, November 2009.
- [3] Y.E. Nesterov. *A Method for Solving the Convex Programming Problem with Convergence Rate  $O(1/(k^2))$* . Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543-547 (Russian).
- [4] Perraudin, N. and Shuman, D. and Puy, G. and Vandergheynst, P; *UNLocBoX A matlab convex optimization toolbox usin proximal splitting methods*. ArXiv e-prints; February, 2014; <http://arxiv-web3.library.cornell.edu/abs/1402.0779>
- [5] Jing M and Wang G. "Convergence of Simultaneous Algebraic Reconstruction Technique (SART)". *IEEE Transaction on Image Processing*. Vol. 12, No. 8, Aug 2003.  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.5402rep=rep1type=pdf>
- [6] Zhu Z, Wahid K, et al. "Improved Compressed Sensing-Based Algorithm for Sparse-View CT Image Reconstruction". *Computational and Mathematical Methods in Medicine*. Vol 2013 Article ID 185750.  
<https://www.hindawi.com/journals/cmmm/2013/185750/>