CS754 Assignment-2

Saksham Rathi, Ekansh Ravi Shankar, Kshitij Vaidya

Declaration: The work submitted is our own, and we have adhered to the principles of academic honesty while completing and submitting this work. We have not referred to any unauthorized sources, and we have not used generative AI tools for the work submitted here.

Question 2

Implementation of ISTA

We implement the Iterative Soft Thresholding Algorithm (ISTA) for solving the LASSO problem which is given by:

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ is the product of the sensing matrix and the DCT basis matrix, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ and $\lambda > 0$ is the regularization parameter. The algorithm also requires us to decide on the step size α and the number of iterations to run the algorithm for. The constant α is taken as the maximum eigenvalue of $A^T A$ and the number of iterations is taken as 1000.

$$\alpha = \max(\text{eig}(A^T A)) \tag{2}$$

Before running the algorithm on the **barbara.png**, we add some Gaussian Noise to the image and also normalise the cells of the image to be in the range $[0, 1]$. We then divide the image into overlapping patches of size 8×8 and run the ISTA algorithm on each patch. We then reconstruct the image from the patches and display the reconstructed image. The results are shown below:

ISTA Continued

The reconstruction does result in a decently accurate recovery of the image. The image is not as sharp as the original image but the major features are still visible. We quantise the quality of the reconstruction by measuring the Root Mean Square Error between the Original Image and the Reconstructed Image. The RMSE value is calculated as :

$$\text{RMSE} = \frac{\|X(:) - \hat{X}(:)\|}{\|X(:)\|_2} \tag{3}$$

From the above formula, the RMSE value calculated for the ISTA reconstruction of the Barbara Image is **0.14027**. This value indicates that the reconstruction is quite accurate and the image is well recovered. While increasing the number of iterations and tuning the regularization parameter λ can improve the quality of the reconstruction, the current reconstruction is good. Next, we replicate the result on the top left 256×256 patch of the **Goldhill Image**. The visual reconstruction of this image is slightly better as evident from the RMSE value of **0.088862**. The

/home/kshitij-vaidya/CS754-Assignments/HW2/assignment2/2/FISTA_Barbara

Figure 1: ISTA Reconstruction of the Barbara Image

image results are shown below.

/home/kshitij-vaidya/CS754-Assignments/HW2/assignment2/2/ISTA_Goldhill

Figure 2: ISTA Reconstruction of the Goldhill Image

Implementation if FISTA

After implementing the ISTA algorithm, we implement the Fast Iterative Soft Thresholding Algorithm (FISTA) for solving the LASSO problem. The FISTA algorithm is an accelerated

version of the ISTA algorithm and is given by the following update rules:

$$x_k = \text{Soft}(y_k - \alpha A^T(Ay_k - b), \lambda\alpha) \tag{4}$$

$$y_{k+1} = x_k + \frac{t_k - 1}{t_k + 2}(x_k - x_{k-1}) \tag{5}$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2} \tag{6}$$

We implement this update rule for the FISTA algorithm and run the algorithm on the same images as before. The results are shown below. The RMSE values for the Barbara and Goldhill images are **0.14065** and **0.087743** respectively. The RMSE values are slightly better than the ISTA algorithm and the visual reconstruction is also better. The images are shown below. This means that the FISTA algorithm is faster than the ISTA algorithm and converges to a better solution in the same number of iterations.

/home/kshitij-vaidya/CS754-Assignments/HW2/assignment2/2/FISTA_Barbara

Figure 3: FISTA Reconstruction of the Barbara Image

Comparison of ISTA and FISTA

0.1 Convergence of ISTA and FISTA

ISTA follows the gradient descent for the soft-thresholding operator. The convergence rate of the ISTA algorithm is sublinear and the decrease in the objective function's value follows the order:

$$F(x_k) = F(x^*) \le \frac{\alpha L(f)\|x_0 - x^*\|^2}{2k} \tag{7}$$

Here α is a constant, L(f) is the Lipschitz constant of the gradient of the objective function and x^* is the optimal solution. The resultant convergence rate is $\mathcal{O}(1/k)$, which is slow and requires a large number of iterations to converge to the optimal solution.

/home/kshitij-vaidya/CS754-Assignments/HW2/assignment2/2/FISTA_Goldhil

Figure 4: FISTA Reconstruction of the Goldhill Image

FISTA, on the other hand accelerates its convergence by incorporating a momentum term. The term is computed based off of both the present and the previous iterates, leading to a faster convergence rate. The global convergece rate of the FISTA algorithm is given by:

$$F(x_k) = F(x^*) \leq \frac{2\|x_0 - x^*\|^2}{(k+1)^2} \tag{8}$$

The convergence rate of the FISTA algorithm is $\mathcal{O}(1/k^2)$, which is faster than the ISTA algorithm. This is why the FISTA algorithm converges faster than the ISTA algorithm and provides a better reconstruction in the same number of iterations.

0.2 Why is FISTA faster than ISTA?

The key reason FISTA outperforms ISTA is because of the momentum like step which combines the information from the previous iterates. This allows the algorithm to take larger and more effective steps towards the optimal solution. The core idea of the algorithm is inspired by Nesterov's optimal gradient method. The update rule for FISTA is:

$$x_k = p_L(y_k), \quad y_{k+1} = x_k + \left(\frac{t_k - 1}{t_k + 2}\right)(x_k - x_{k-1}) \tag{9}$$

Here p_L is the proximal operator associated with the objective function and t_k is the momentum term. The momentum term is updated as:

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

The convergence of FISTA is established by proving that the function approaches the optimal value at a rate of $\mathcal{O}(1/k^2)$. This is why FISTA is faster than ISTA and provides a better reconstruction in the same number of iterations.

Lemma : Recursion in FISTA : The sequence of values generated by the FISTA algorithm satisfy the following inequality:

$$\frac{2}{L}\left(t_k^2 F(x_k) - t_{k+1}^2 F(x_{k+1})\right) \geq \|u_{k+1}\|^2 - \|u_k\|^2 \tag{11}$$

where $u_k = t_k x_k - t_{k-1} x_{k-1} - x^*$. This lemma and the smoothness of the objective function are used to prove the convergence of the FISTA algorithm. While both problems use similar update rules, the momentum term in FISTA makes it a much faster algorithm especially for large scale problems.