Assignment 3: CS 754, Advanced Image Processing

Shubham Kar
180070058

Garaga VVS Krishna Vamsi
180070020

March 2022

1. Your task here is to implement the ISTA algorithm for the following three cases:

   (a) Consider the image from the homework folder. Add iid Gaussian noise of mean 0 and variance 3 (on a [0,255] scale) to it, using the 'randn' function in MATLAB. Thus $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\eta}$ where $\boldsymbol{\eta} \sim \mathcal{N}(0, 3)$ (earlier the variance was mistakenly marked as 4). You should obtain $\boldsymbol{x}$ from $\boldsymbol{y}$ using the fact that patches from $\boldsymbol{x}$ have a sparse or near-sparse representation in the 2D-DCT basis.

   (b) Divide the image shared in the homework folder into patches of size 8×8. Let $\boldsymbol{x_i}$ be the vectorized version of the i^{th} patch. Consider the measurement $\boldsymbol{y_i} = \boldsymbol{\Phi x_i}$ where $\boldsymbol{\Phi}$ is a 32×64 matrix with entries drawn iid from $\mathcal{N}(0, 1)$. Note that $\boldsymbol{x_i}$ has a near-sparse representation in the 2D-DCT basis $\boldsymbol{U}$ which is computed in MATLAB as 'kron(dctmtx(8)',dctmtx(8)')'. In other words, $\boldsymbol{x_i} = \boldsymbol{U\theta_i}$ where $\boldsymbol{\theta_i}$ is a near-sparse vector. Your job is to reconstruct each $\boldsymbol{x_i}$ given $\boldsymbol{y_i}$ and $\boldsymbol{\Phi}$ using ISTA. Then you should reconstruct the image by averaging the overlapping patches. You should choose the α parameter in the ISTA algorithm judiciously. Choose $\lambda = 1$ (for a [0,255] image). Display the reconstructed image in your report. State the RMSE given as $\|X(:) - \hat{X}(:)\|_2 / \|X(:)\|_2$ where $\hat{X}$ is the reconstructed image and X is the true image. [15 points]

Sol:

(a) The measurement matrix($\boldsymbol{\Phi}$) here is the simple sampling matrix and therefore the coherence of the measurement matrix with the DCT matrix is well within the limit. Therefore, we can use compressed sensing here to retrieve **x**.
Noise added to the image is compared with the original image in Figure 1. The RMSE was found out to be 0.0119. The reconstructed image is shown in Figure 2.

(b) The same algorithm has to be applied for this part of the question as well except for the fact that instead of the measurement matrix being the sampling matrix, it's a matrix with standard Gaussian variables. Also, the number of measurements are less than the number of input pixels. The RMSE error comes out to be 0.03781. The reconstructed image is compared with the original image in Figure 3.

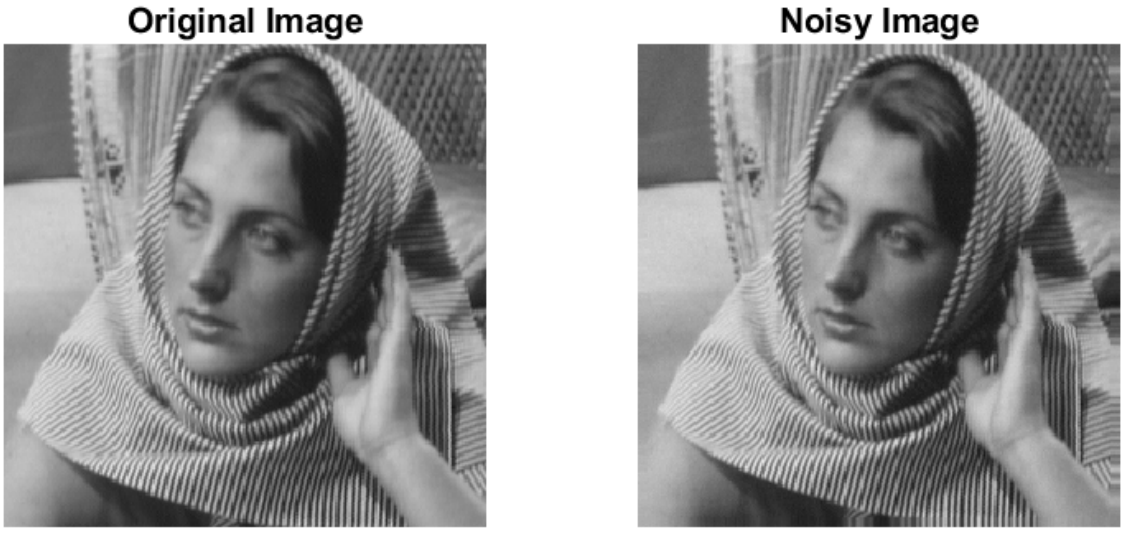


Figure 1: Comparison of the noisy image with the original image

Figure 2: Reconstruction of the image using patch-wise analysis with DCT coefficients of the patches

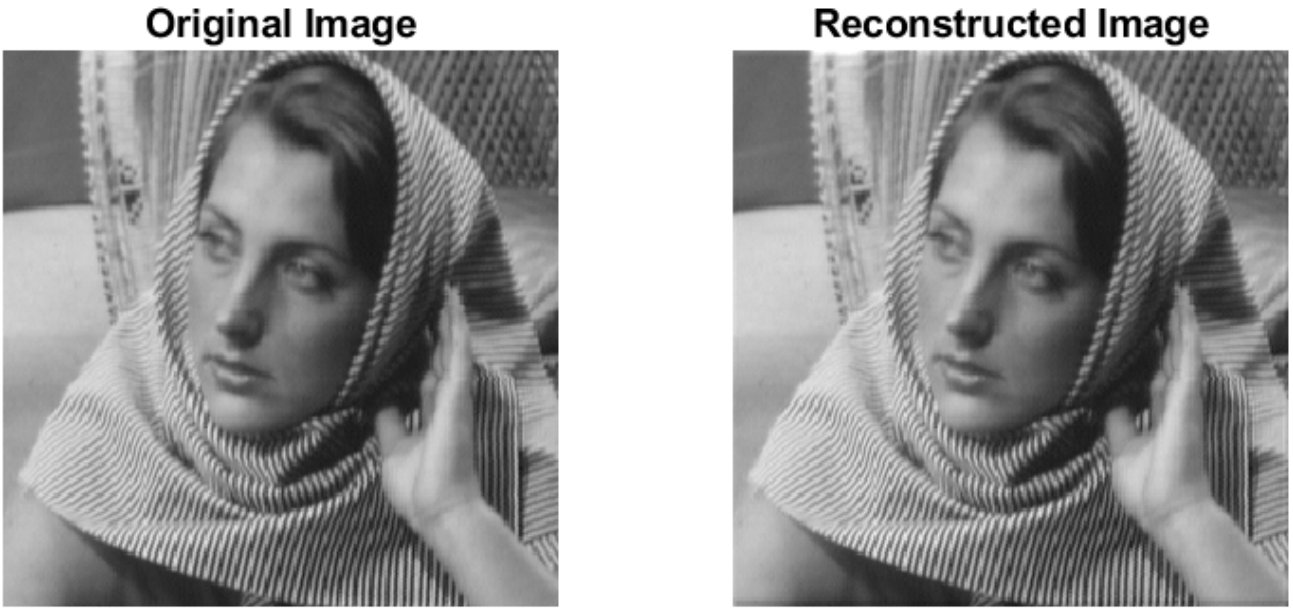


Figure 3: Comparison of the original image with the reconstructed image

2. Download the book 'Statistical Learning with Sparsity: The Lasso and Generalizations' from `https://web.stanford.edu/~hastie/StatLearnSparsity_files/SLS_corrected_1.4.16.pdf`, which is the website of one of the authors. (The book can be officially downloaded from this online source). Your task is to trace through the steps of the proof of Theorem 11.1(b). This theorem essentially derives error bounds on the minimum of the following objective function: $J(\boldsymbol{\beta}) = \frac{1}{2N}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda_N\|\boldsymbol{\beta}\|_1$ where λ_N is a regularization parameter, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the unknown sparse signal, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{w}$ is a measurement vector with N values, $\boldsymbol{w}$ is a zero-mean i.i.d. Gaussian noise vector whose each element has standard deviation σ and $\boldsymbol{X} \in \mathbb{R}^{N\times p}$ is a sensing matrix whose every column is unit normalized. This particular estimator (i.e. minimizer of $J(\boldsymbol{x})$ for $\boldsymbol{x}$) is called the LASSO in the statistics literature. The theorem derives a statistical bound on λ also. Your task is split up in the following manner:

   (a) Define the restricted eigenvalue condition (the answer's there in the book and you are allowed to read it, but you also need to understand it).

   (b) Starting from equation 11.20 on page 309 - explain why $G(\hat{v}) \leq G(0)$.

   (c) Do the algebra to obtain equation 11.21.

   (d) Do the algebra in more detail to obtain equation 11.22 (state the exact method of application of Holder's inequality - check the wiki article on it, if you want to find out what this inequality states).

   (e) Derive equation 11.23.

   (f) Assuming Lemma 11.1 is true and now that you have derived equation 11.23, complete the proof for the final error bound for equation 11.14b.

   (g) In which part of the proof does the bound $\lambda_N \geq 2\frac{\|\boldsymbol{X}^T\boldsymbol{w}\|_\infty}{N}$ show up? Explain.

   (h) Why is the cone constraint required? You may read the rest of the chapter to find the answer.

   (i) Read example 11.1 which tells you how to put a tail bound on λ_N assuming that the noise vector $\boldsymbol{w}$ is zero-mean Gaussian with standard deviation σ. Given this, state the advantages of this theorem over Theorem 3 that we did in class. You may read parts of the rest of the chapter to answer this question. What are the advantages of Theorem 3 over this particular theorem?

   (j) Now read Theorem 1.10 till corollary 1.2 and comments on it concerning an estimator called the 'Dantzig selector', in the tutorial 'Introduction to Compressed Sensing' by Davenport, Duarte, Eldar and Kuttyniok. You can find it here: `http://www.ecs.umass.edu/~mduarte/images/IntroCS.pdf` or at `https://webee.technion.ac.il/Sites/People/YoninaEldar/files/ddek.pdf`. What is the common thread between the bounds on the 'Dantzig selector' and the LASSO? [2 x 8 + 4 + 4 = 24 points]

Sol:

(a) Consider the loss function for Regression

$$f_N(\beta) = \frac{1}{2N}\|y - X\beta\|_2^2$$

One can clearly see that is function is always convex. The double derivative w.r.t to β gives,

$$\nabla^2 f_N(\beta) = X^T X/N \qquad \forall\beta$$

Now the additional condition that we need for it to be strongly convex is that all the eigenvalues of the p×p positive semi-definite matrix $X^T X$ should be uniformly bounded away from zero. However, we know that the rank of $X^T X$ can be atmost min(N,p). Hence in the case of p > N, the matrix $X^T X$ is rank-deficient and hence the loss function is not strongly convex. For this reason we relax our notion of strong convexity condition, We make it such that it is only necessary to impose the type of strong

convexity for some subset $C \subset R^p$ of possible perturbation vectors v. So, we say that a function f satisfies restricted strong convexity at β^* with respect to C if there is a constant $\gamma > 0$ such that,

$$\frac{v^T \nabla^2 f(\beta) v}{\|v\|_2^2} \geq \gamma \text{ for all nonzero } v \in C$$

For the case of regression loss this is equivalent to lower bounding the **restricted eigenvalues** of the model

$$\frac{v^T X^T X v}{N\|v\|_2^2} \geq \gamma \text{ for all nonzero } v \in C$$

This is called the restricted eigenvalues condition.

(b) We know that $\hat{\beta}$ is the solution of the LASSO function. Hence $\hat{\beta}$ minimizes,

$$f(\beta) = \frac{1}{2N}\|y - X\beta\|_2^2 + \lambda_N \|\beta\|_1$$

we obtain G(v) from f(β) by substituting $\beta = \beta^* + v$. Hence G(v) will be maximised when $v + \beta^* = \hat{\beta}$, that implies $\hat{v} = \hat{\beta} - \beta^*$ minimizes

$$G(v) = \frac{1}{2N}\|y - X(\beta^* + v)\|_2^2 + \lambda_N \|\beta^* + v\|_1$$
$$\text{Hence,} G(\hat{v}) \leq G(v) \qquad \forall v$$

Hence G($\hat{v}$) $\leq$ G(0).

(c) Using G($\hat{v}$) $\leq$ G(0),

$$\frac{1}{2N}\|y - X(\beta^* + \hat{v})\|_2^2 + \lambda_N \|\beta^* + \hat{v}\|_1 \leq \frac{1}{2N}\|y - X(\beta^* + 0)\|_2^2 + \lambda_N \|\beta^* + 0\|_1$$
$$\frac{1}{2N}\|y - X(\beta^* + \hat{v})\|_2^2 + \lambda_N \|\beta^* + \hat{v}\|_1 \leq \frac{1}{2N}\|y - X\beta^*\|_2^2 + \lambda_N \|\beta^*\|_1$$
Using y $= X\beta^* + w$
$$\frac{1}{2N}\|w - X\hat{v}\|_2^2 + \lambda_N \|\beta^* + \hat{v}\|_1 \leq \frac{1}{2N}\|w\|_2^2 + \lambda_N \|\beta^*\|_1$$
$$\frac{1}{2N}\|w - X\hat{v}\|_2^2 - \frac{1}{2N}\|w\|_2^2 \leq \lambda_N \|\beta^*\|_1 - \lambda_N \|\beta^* + \hat{v}\|_1$$
$$\frac{1}{2N}(\|w\|_2^2 + \|X\hat{v}\|_2^2 - w^T X\hat{v} - (X\hat{v})^T w - \|w\|_2^2) \leq \lambda_N(\|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1)$$
As $w^T X\hat{v}$ is 1×1 matrix,
$$\frac{1}{2N}(\|X\hat{v}\|_2^2 - 2w^T X\hat{v}) \leq \lambda_N(\|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1)$$
$$\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{w^T X\hat{v}}{N} + \lambda_N(\|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1) \qquad (11.21)$$

(d) Consider the scalar $w^T X\hat{v}$ as the product of a 1×p matrix $w^T X$ and p×1 matrix $\hat{v}$. Hence this can also be written as dot-product of two p×1 vectors v and $(w^T X)^T = X^T w$. Hence using holder's inequality, we can write

$$w^T X\hat{v} = \sum_{i=1}^{p} \hat{v}_i (X^T w)_i \leq \sum_{i=1}^{p} |\hat{v}_i (X^T w)_i| = \sum_{i=1}^{p} |\hat{v}_i||(X^T w)_i| \leq \sum_{i=1}^{p} |\hat{v}_i| max_i(|(X^T w)_i|)$$
$$w^T X\hat{v} \leq max_i(|(X^T w)_i|) \sum_{i=1}^{p} |\hat{v}_i| = \|X^T w\|_\infty \|\hat{v}\|_1 \qquad (1)$$

using the fact that $\beta_{S^c} = 0$ as it is an S-sparse vector.

$$\|\beta^* + \hat{v}\|_1 = \|\beta^*_S + 0 + \hat{v}_S + \hat{v}_{S^c}\|_1 = \|\beta^*_S + \hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1$$

because S and S^c are completely disjoint. Now using the reverse triangle inequality

$$\|\beta^* + \hat{v}\|_1 \geq \|\beta^*_S\|_1 - \|\hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1$$

$$\|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1 \leq \|\beta^*\|_1 - \|\beta^*_S\|_1 + \|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1$$

$$\text{As, } \|\beta^*\|_1 = \|\beta^*_S\|_1$$

$$\|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1 \leq \|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1 \tag{2}$$

Using the inequalities (1) and (2) on (11.21), we will get,

$$\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{\|X^T w\|_\infty \|\hat{v}\|_1}{N} + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1) \tag{11.22}$$

(e) By assumption, $\frac{\|X^T w\|_\infty}{N} \leq \frac{\lambda_N}{2}$. Substituting in (11.22), we get

$$\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{\lambda_N\|\hat{v}\|_1}{2} + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1) = \frac{\lambda_N\|\hat{v}_S + \hat{v}_{S^c}\|_1}{2} + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1)$$

$$\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{\lambda_N(\|\hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1)}{2} + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1) = \lambda_N(\frac{3\|\hat{v}_S\|_1}{2} - \frac{\|\hat{v}_{S^c}\|_1}{2})$$

$$\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{3\lambda_N\|\hat{v}_S\|_1}{2} \leq \frac{3\lambda_N\sqrt{k}\|\hat{v}\|_2}{2}, \tag{11.23}$$

Here, k is the subset of size S

The last step ($\|\hat{v}_S\|_1 \leq \sqrt{k}\|\hat{v}\|_2$) is done using Cauchy-schwartz inequality as follows,

$$\|\hat{v}_S\|_1 = \|I_S.\hat{v}_S\| = \|I_S.\hat{v}\|_2 \leq \|I_S\|_2\|\hat{v}\|_2 = \sqrt{k}\|\hat{v}|_2$$

$$\text{As, } I_S.\hat{v} \text{ is a scalar, } \|I_S.\hat{v}\| = \|I_S.\hat{v}\|_2$$

where I_S is the indicator that the current index is in subset S.

i.e., it is a vector k 1's and zeros elsewhere. That implies $\|I_S\|_2 = \sqrt{k}$

(f) Assuming Lemma 11.1, the error $\hat{v}$ belongs to cone set C and we can use the Restricted eigenvalues condition for $\hat{v}$

$$\frac{\hat{v}^T X^T X\hat{v}}{N\|\hat{v}\|_2^2} = \frac{\|X\hat{v}\|_2^2}{N\|\hat{v}\|_2^2} \geq \gamma$$

$$\frac{\gamma\|\hat{v}\|_2^2}{2} \leq \frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{3\lambda_N\sqrt{k}\|\hat{v}\|_2}{2} \qquad \text{(from 11.23)}$$

$$\|\hat{v}\|_2 = \|\hat{\beta} - \beta^*\|_2 \leq \frac{3\lambda_N\sqrt{k}}{\gamma} \qquad (\hat{v} = \hat{\beta} - \beta^*, \text{ from part } (b))$$

$$\|\hat{\beta} - \beta^*\|_2 \leq \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\lambda_N \tag{11.14b}$$

(g) The inequality $\lambda_N \geq 2\frac{\|\boldsymbol{X}^T\boldsymbol{w}\|_\infty}{N}$ is used at two places in the proof, Firstly, it is used in part (e), i.e., to get (11.23) from (11.22). This gives us $\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{3\lambda_N\|\hat{v}_S\|_1}{2} - \frac{\lambda_N\|\hat{v}_{S^c}\|_1}{2}$. Secondly it is used in the final step to prove Lemma 11.1 as assuming this bound, will give us the cone constraint ($\|\hat{v}_{S^c}\|_1 \leq \alpha\|\hat{v}_S\|_1$). Seeing that the LHS of the last equation is always greater than 0,

$$\frac{3\lambda_N\|\hat{v}_S\|_1}{2} - \frac{\lambda_N\|\hat{v}_{S^c}\|_1}{2} \geq 0$$

$$\|\hat{v}_{S^c}\|_1 \leq 3\|\hat{v}_S\|_1 (\text{ cone constraint})$$

(h) At first we say we need strictly convex condition to minimize the loss next we acknowledge that is not always possible in practice hence we relaxed the condition to restricted eigen values condition

$$\frac{v^T X^T X v}{N\|v\|_2^2} \geq \gamma \text{ for all nonzero } v \in C$$

Now what the relevant constraint set C. For appropriate choices of the L_1-ball radius or equivalently, of the regularization parameter λ_N, it turns out that the lasso error $\|\hat{v}\| = \|\hat{\beta} - \beta^*\|$ satisfies a cone constraint. Hence once it satisfies the cone constraint, we can use the restricted eigen values condition and do the final step of the above proof.

(i) Theorem 3 states that if the matrix A satisfies the RIP of order 2S with $\delta_{2S} < 0.41$ and $\hat{\beta}$ is the solution obtained by solving the problem

$$min\|\beta\|_1 \text{ s.t. } \|Y - A\beta\|_2 < \epsilon$$

and Y = Aβ^* + w,

$$\|\hat{\beta} - \beta^*\| \leq \frac{C_0}{\sqrt{S}}\|\beta^* - \beta_S^*\| + C_0\sigma$$

Where σ^2 is the variance of w

Since, we take β^* to be S-sparse

$$\|\hat{\beta} - \beta^*\| \leq C_0\sigma$$

Theorem 11.1(b) states that if $\lambda_N \geq \frac{2\|X\|_\infty}{N}$ then the solution $\hat{\beta}$ of the LASSO function satisfies the bound,

$$\|\hat{\beta} - \beta^*\| \leq \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\lambda_N \tag{1}$$

According to example 11.1, if we choose $\lambda_N = 2\sigma\sqrt{\tau\frac{log(p)}{N}}$, then the bound obtained using theorem 11.1(b) is satisfied with high probability. i.e,

$$\|\hat{\beta} - \beta^*\| \leq \frac{6\sigma}{\gamma}\sqrt{\frac{k\tau log(p)}{N}} \tag{2}$$

Benefits of Theorem 11.1(b) over Theorem 3:

- We can see that the error bound decreases as we increase the number of measurements significantly. Of course the bound given by theorem3 will also be affected indirectly as δ_{2S} changes.
- In theorem 3 we need to evaluate the RIC of the matrix which will need to find Eigenvalues of different subset matrices of A, but here the search is limited to the cone set and does not need the explicit calculation of RIC.
- The rate achieved by this theorem, including the logarithmic factor is known to be optimal and cannot be substantially improved upon by any estimator.

Benefits of Theorem 3 over Theorem 11.1(b):

- This theorem doesn't consider compressible signals (that is when β^* is not S-sparse) but in many cases the signals are not exactly sparse but have very small values for some extra elements. These cases are however handled by Theorem3

(j) Both the Dantzig selector and the LASSO bounds are proportional to $\sqrt{klog(p)}$. Hence the bounds reduce if the sparsity of the solution increases (less number of non-zero entries). Even in the proof, both methods account for the interaction between the dictionary and the noise vector on the error bound, that is they bound their maximum correlation $\|X^T w\|_\infty$ (in this theorem), $\|A^T e\|_\infty$ (in Dantzig selector) by some constant.

3. In this task, you will you use the well-known package L1_LS from `https://stanford.edu/~boyd/l1_ls/`. This package is often used for compressed sensing solution, but here you will use it for the purpose of tomographic reconstruction. The homework folder contains images of two slices taken from an MR volume of the brain. Create measurements by parallel beam tomographic projections at any 18 randomly angles chosen from a uniform distribution on $[0, \pi)$. Use the MATLAB function 'radon' for this purpose. Now perform tomographic reconstruction using the following method: (a) filtered back-projection using the Ram-Lak filter, as implemented in the 'iradon' function in MATLAB, (b) independent CS-based reconstruction for each slice by solving an optimization problem of the form $J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1$, (c) a coupled CS-based reconstruction that takes into account the similarity of the two slices using the model given in the lectures notes on tomography. For parts (b) and (c), use the aforementioned package from Stanford. For part (c), make sure you use a different random set of 18 angles for each of the two slices. The tricky part is careful creation of the forward model matrix $\boldsymbol{A}$ or a function handle representing that matrix, as well as the corresponding adjoint operator $\boldsymbol{A}^T$. Use the 2D-DCT basis for the image representation. Modify the objective function from the lecture notes for the case of three similar slices. Carefully define all terms in the equation but do not re-implement it. For ease of implementation, use square images. For this zero-pad the original images to make them square-shaped before getting the radon projections. You can also specify the output size in the iradon function. You may work with uniformly spaced angles instead of randomly generated angles as the former can give better results. [3+7+8+7 = 25 points]

Sol:

The brain slice images have been padded up to a size of 255×255 for ease of computation with regard to using radon and DCT function handles. The padded images are compared with the original brain slices in Figure 4.

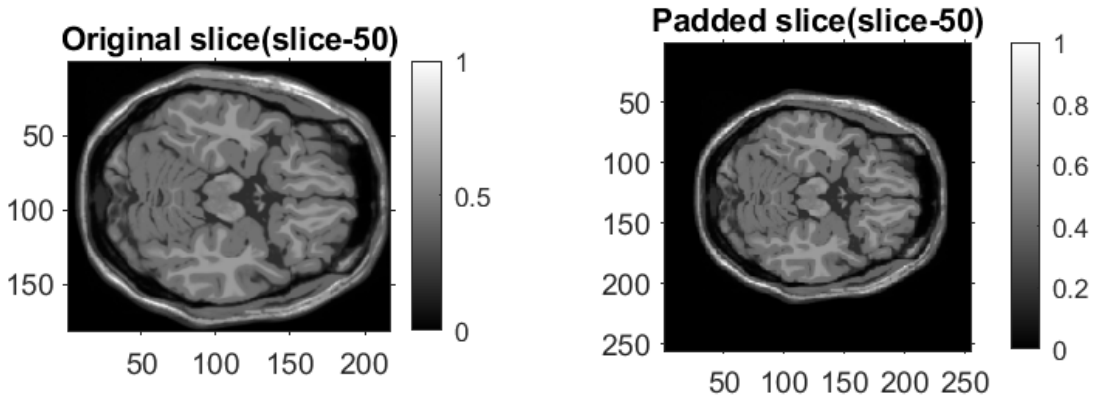


Figure 4: Comparison of original and padded brain slice images

(a) The filtered-back reconstructions for all the brain slices for randomly picked angles and uniformly spaced angles are shown in Figures 5 and 6.

(b) CS-based reconstructions using independent brain slice projections are shown in Figures 7 and 8.

(c) Coupled CS-based reconstruction using two consecutive slices are shown in Figures 9 and 10.

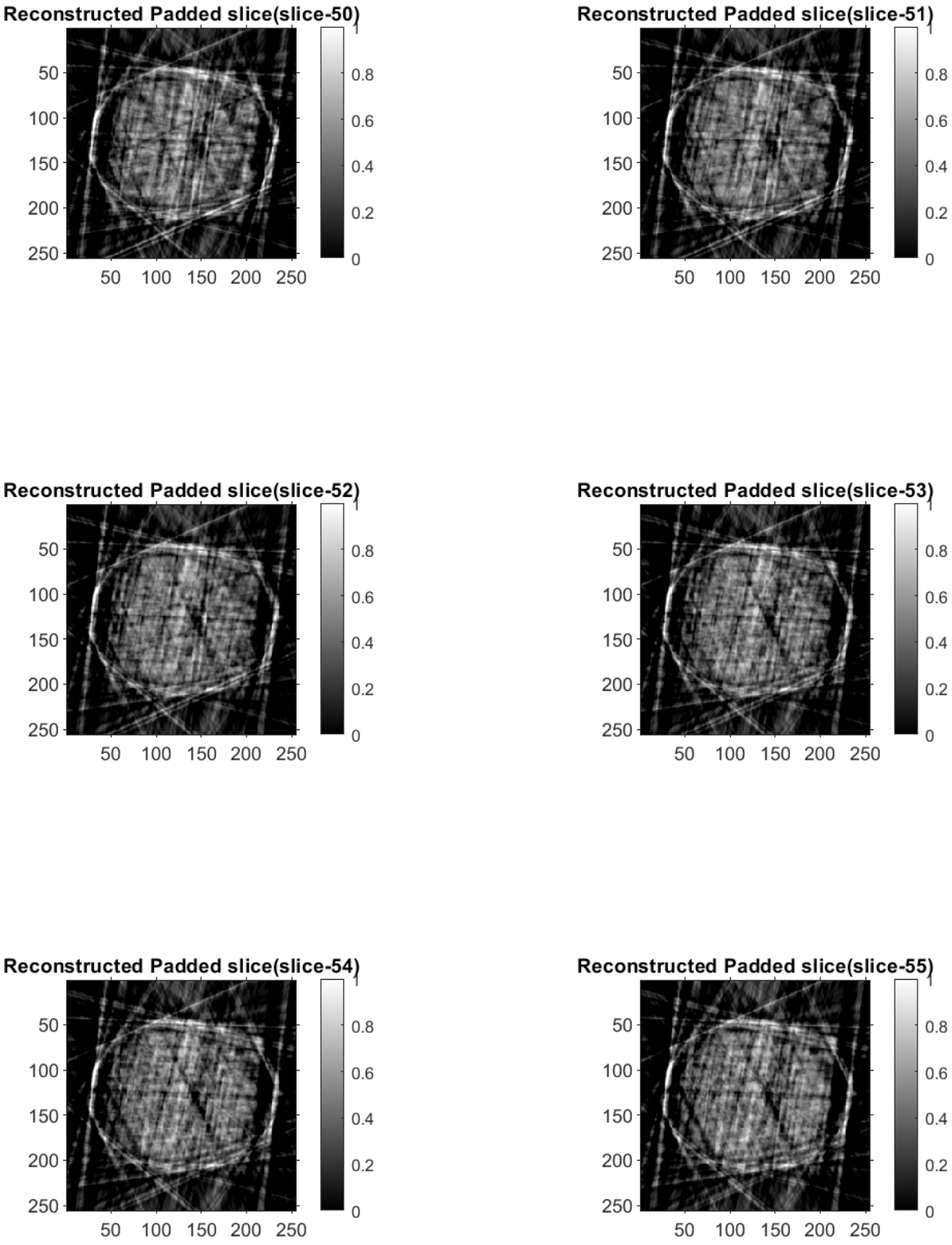


Figure 5: Filtered Back Projections of brain slices using 18 random angles for each slice

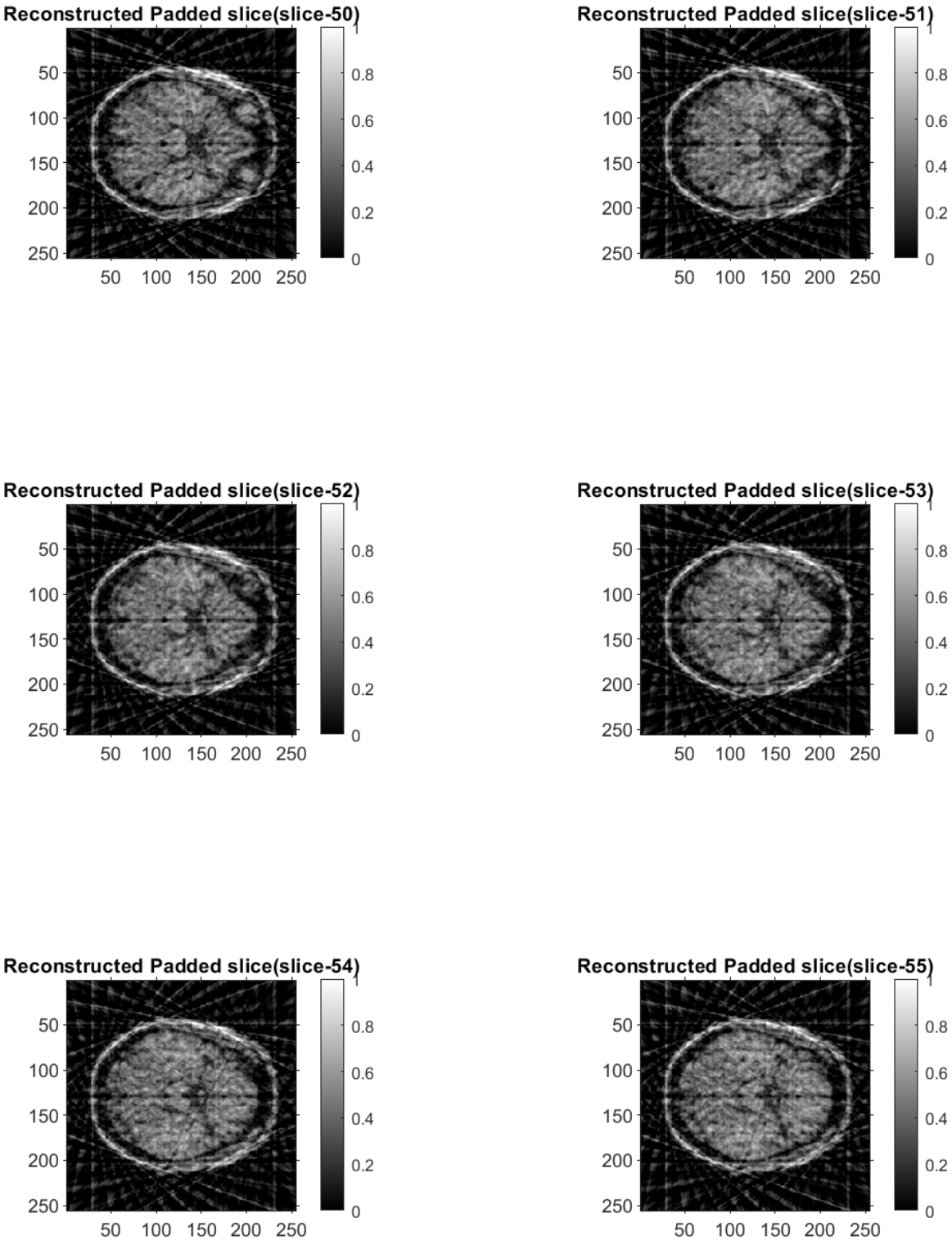


Figure 6: Filtered Back Projections of brain slices using 18 uniformly spaced angles for each slice

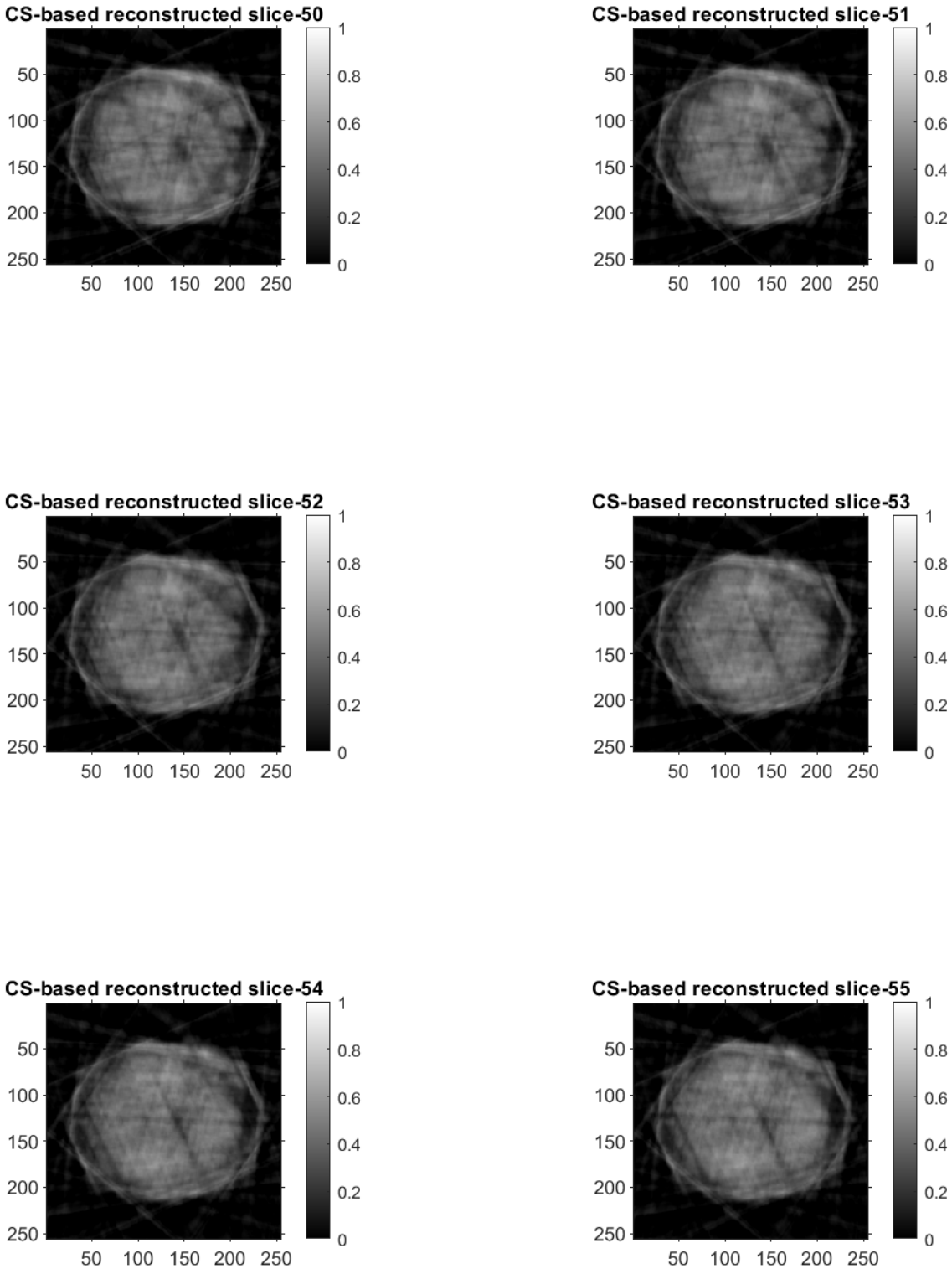


Figure 7: Independent CS-based reconstruction using 18 random angle projections for each slice

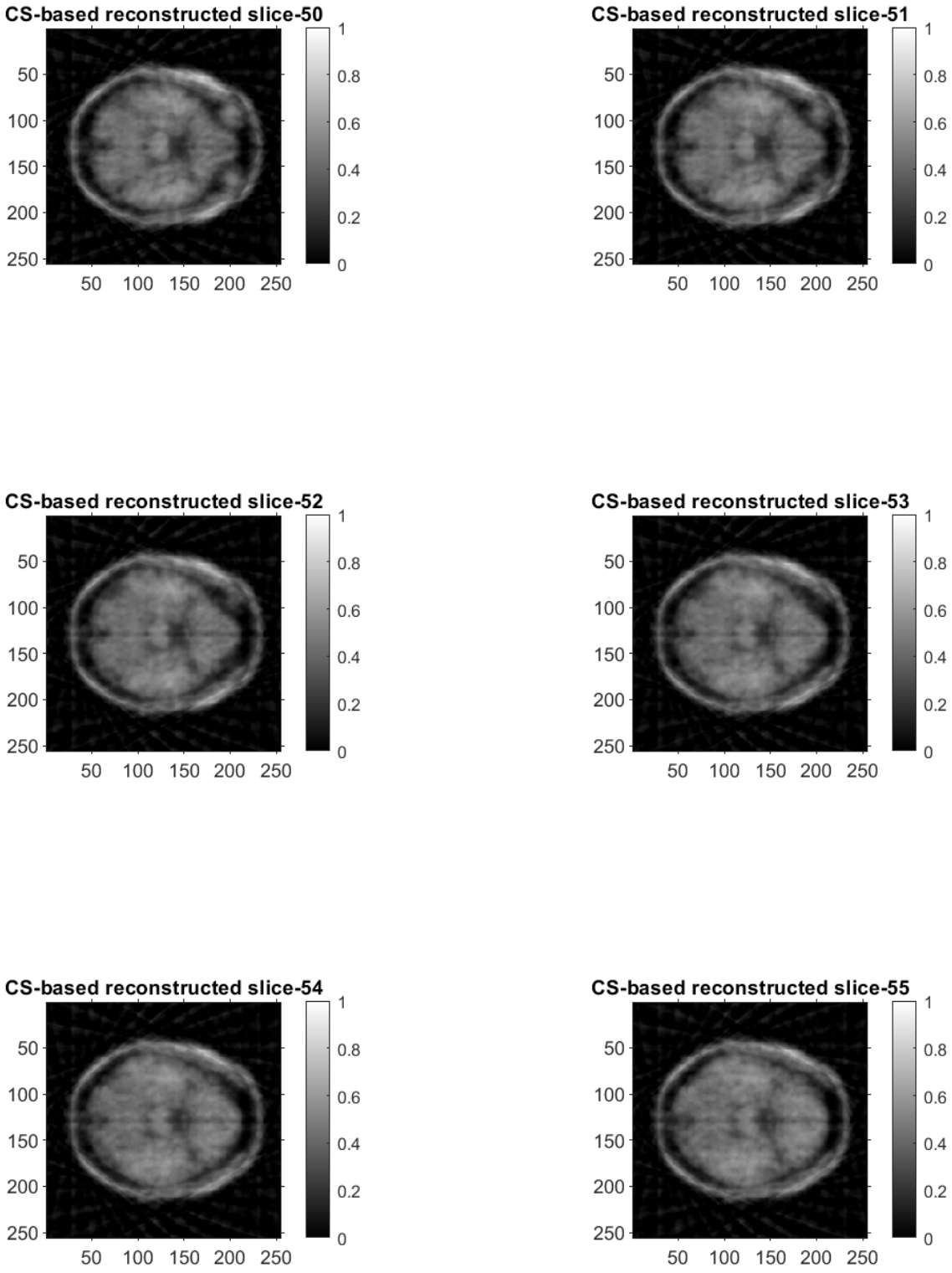


Figure 8: Independent CS-based reconstruction using 18 uniformly spaced angle projections for each slice

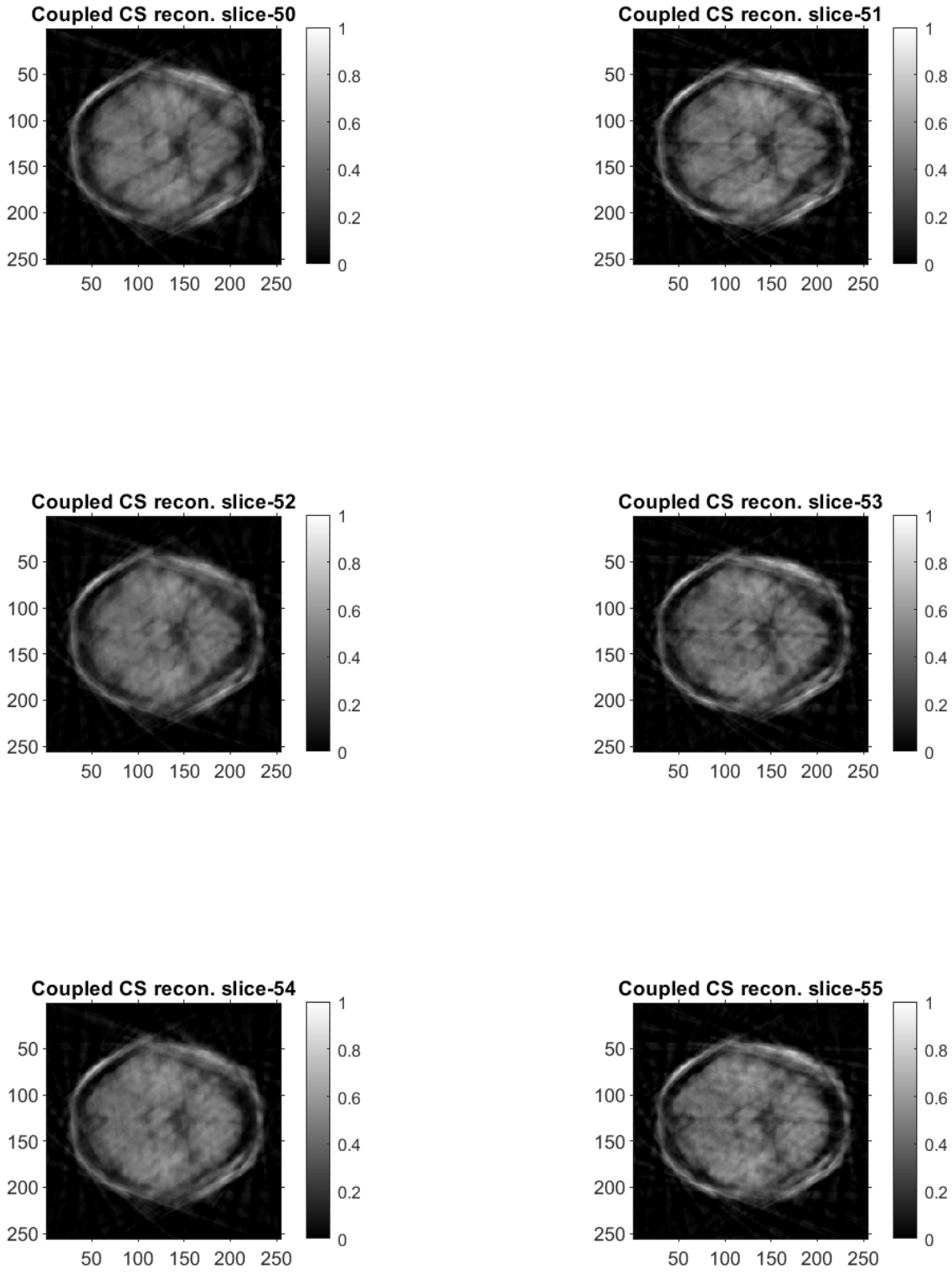


Figure 9: Coupled CS-based reconstruction using 18 **different** random angle projections for each slice

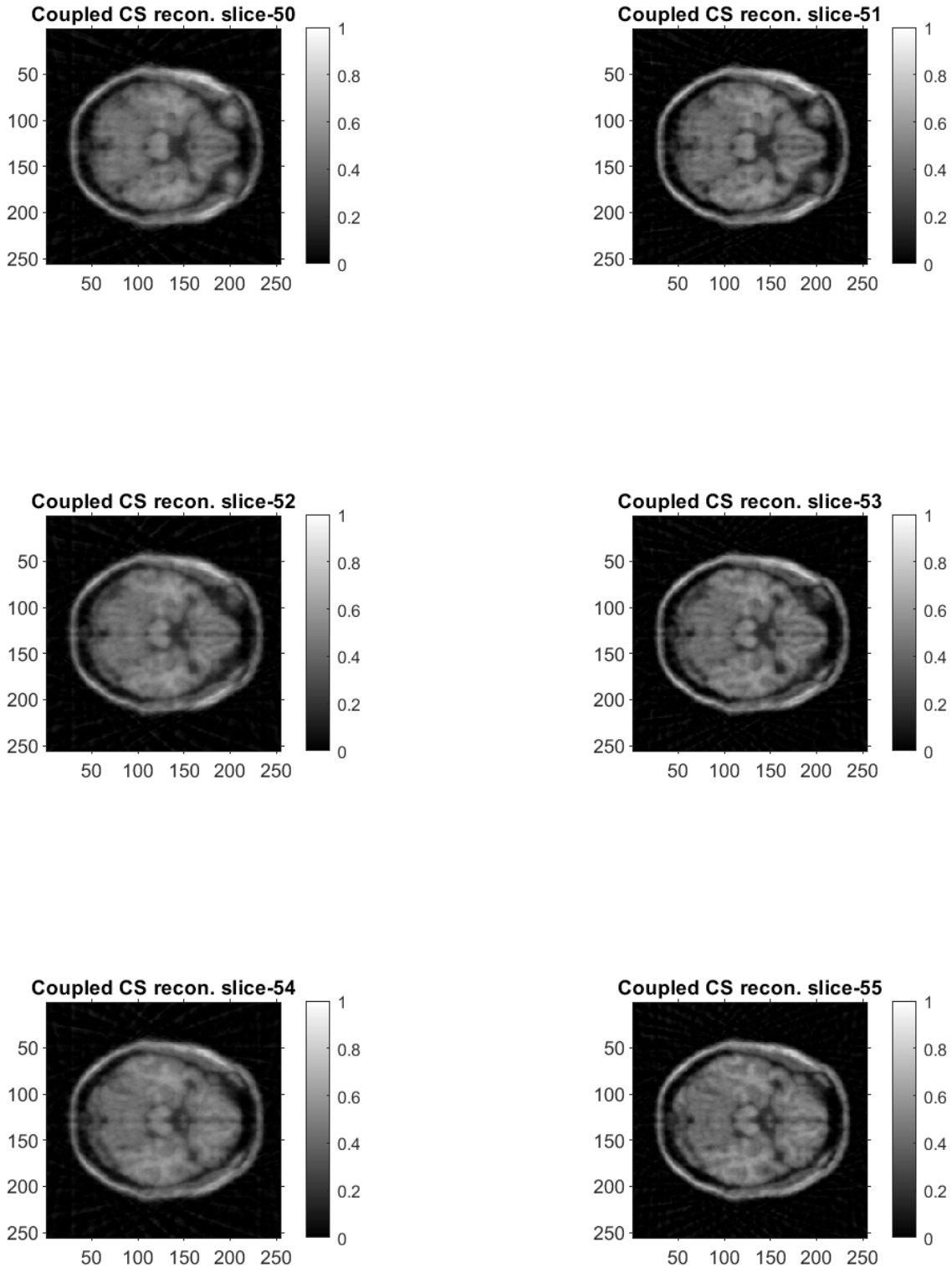


Figure 10: Coupled CS-based reconstruction using 18 **different** uniformly spaced angle projections for each slice

(d) For 3 slice-coupled CS-based reconstruction, we have to modify the objective function accordingly. The new objective function can be written as:

$$E(\beta_1, \beta_2, \beta_3) \;=\; ||\mathbf{y_1} - \mathbf{R_1 U}\beta_1||^2 \;+\; ||\mathbf{y_2} - \mathbf{R_2 U}\beta_2||^2 \;+\; ||\mathbf{y_3} - \mathbf{R_3 U}\beta_3||^2 + \lambda||\beta_1||_1 \;+\; \lambda||\beta_2 - \beta_1||_1 \;+\; \lambda||\beta_3 - \beta_1||_1 \tag{3}$$

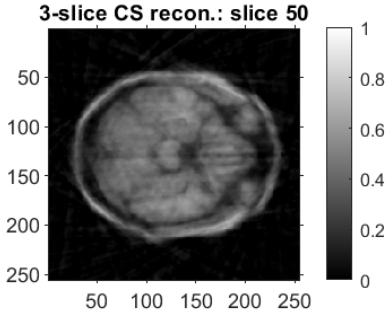
$$\implies E(\beta_1, \beta_2, \beta_3) \;=\; ||\mathbf{y_1} - \mathbf{R_1 U}\beta_1||^2 \;+\; ||\mathbf{y_2} - \mathbf{R_2 U}(\beta_1 + \mathbf{\Delta}\beta_1)||^2 \;+\; ||\mathbf{y_3} - \mathbf{R_3 U}(\beta_1 + \mathbf{\Delta}\beta_2)||^2 + \lambda||\beta_1||_1 \;+\; \lambda||\mathbf{\Delta}\beta_1||_1 \;+\; \lambda||\mathbf{\Delta}\beta_2||_1 \tag{4}$$

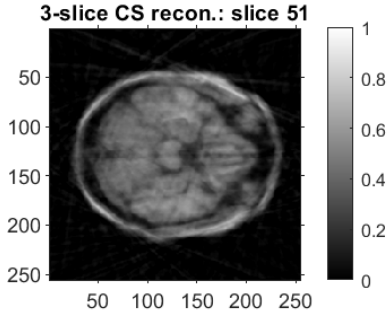
$$\implies E(\beta_1, \beta_2, \beta_3) \;=\; \left|\left|\begin{pmatrix}\mathbf{y_1}\\ \mathbf{y_2}\\ \mathbf{y_3}\end{pmatrix} - \begin{pmatrix}\mathbf{R_1 U} & 0 & 0\\ \mathbf{R_2 U} & \mathbf{R_2 U} & 0\\ \mathbf{R_3 U} & 0 & \mathbf{R_3 U}\end{pmatrix}\begin{pmatrix}\beta_1\\ \mathbf{\Delta}\beta_1\\ \mathbf{\Delta}\beta_2\end{pmatrix}\right|\right| \;+\; \lambda\left|\left|\begin{pmatrix}\beta_1\\ \mathbf{\Delta}\beta_1\\ \mathbf{\Delta}\beta_2\end{pmatrix}\right|\right|_1 \tag{5}$$

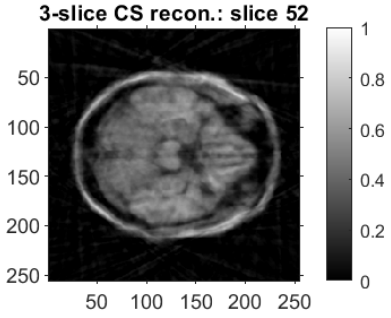
where

- $\mathbf{y_1}$, $\mathbf{y_2}$, and $\mathbf{y_3}$ are the radon transforms of the 3 coupled slices
- β_1, $(\beta_1 + \mathbf{\Delta}\beta_1)$, and $(\beta_1 + \mathbf{\Delta}\beta_2)$ are the 2D-DCT coefficients of the 3 coupled slices respectively, vectorized
- $\mathbf{R_1}$, $\mathbf{R_2}$, and $\mathbf{R_3}$ are the forward model of the radon transforms for each of the 3 slices suing their corresponding tomographic projections
- $\mathbf{U}$ is the 2D-DCT matrix

The 3-slice coupled CS-based reconstruction results are shown in Figures 11, 12, 13, and 14.







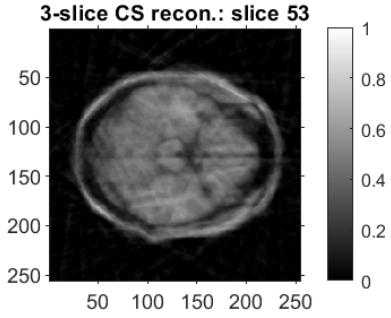
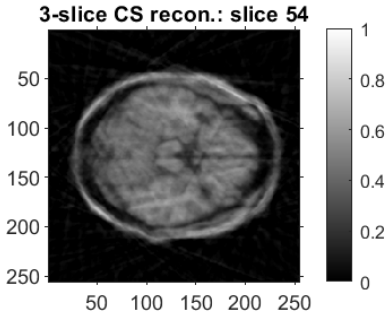


Figure 11: 3 slice-coupled CS-based reconstruction using 18 random angular projections



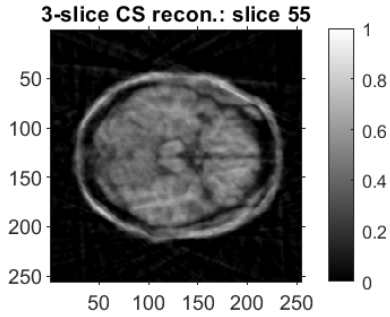
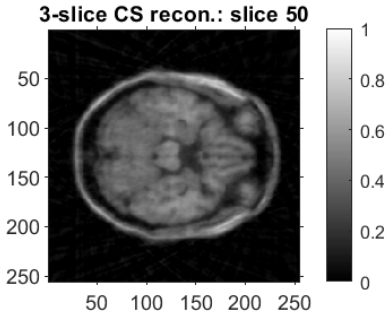
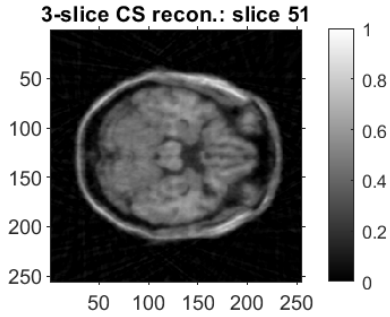
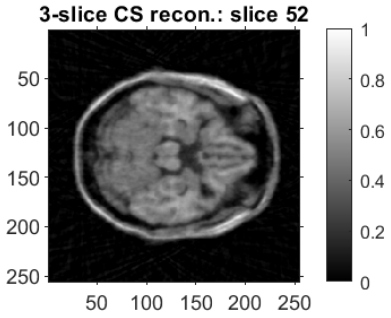


Figure 12: 3 slice-coupled CS-based reconstruction using 18 random angular projections(continued)







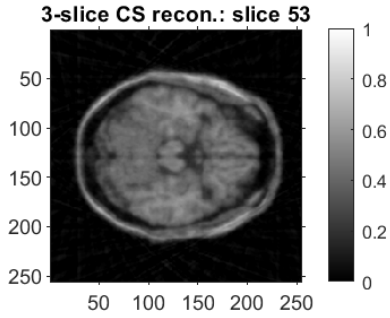


Figure 13: 3 slice-coupled CS-based reconstruction using 18 uniformly spaced angular projections

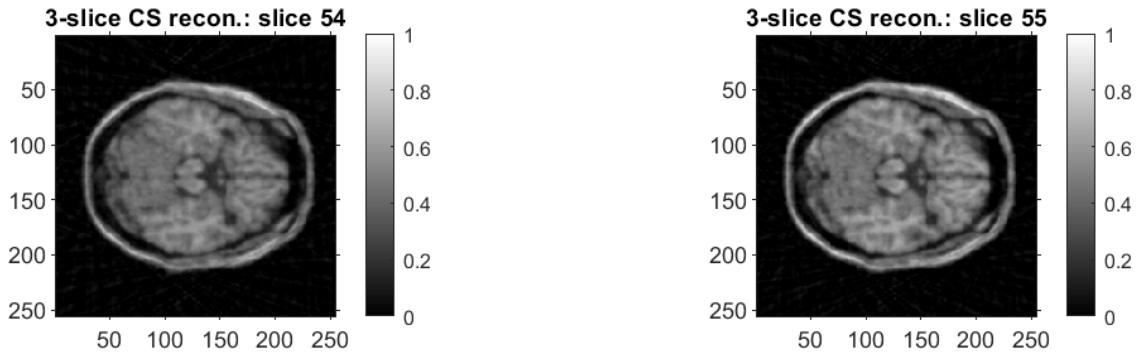


Figure 14: 3 slice-coupled CS-based reconstruction using 18 uniformly spaced angular projections(continued)

4. Here is our Google search question again. You know of the applications of tomography in medicine (CT scanning) and virology/structural biology. Your job is to search for a journal paper from any other field which requires the use of tomographic reconstruction (examples: seismology, agriculture, gemology). State the title, venue and year of publication of the paper. State the mathematical problem defined in the paper. Take care to explain the meaning of all key terms clearly. State the method of optimization that the paper uses to solve the problem. [16 points]

Sol:

(a) **Title:** Compton Scattering Tomography for Agricultural Measurements
Venue: Engenharia Agrícola
Year of Publication: 2006

(b) The paper uses the dependence of the number of scattered photons on soil density and moisture. The experimental setup is as follows:

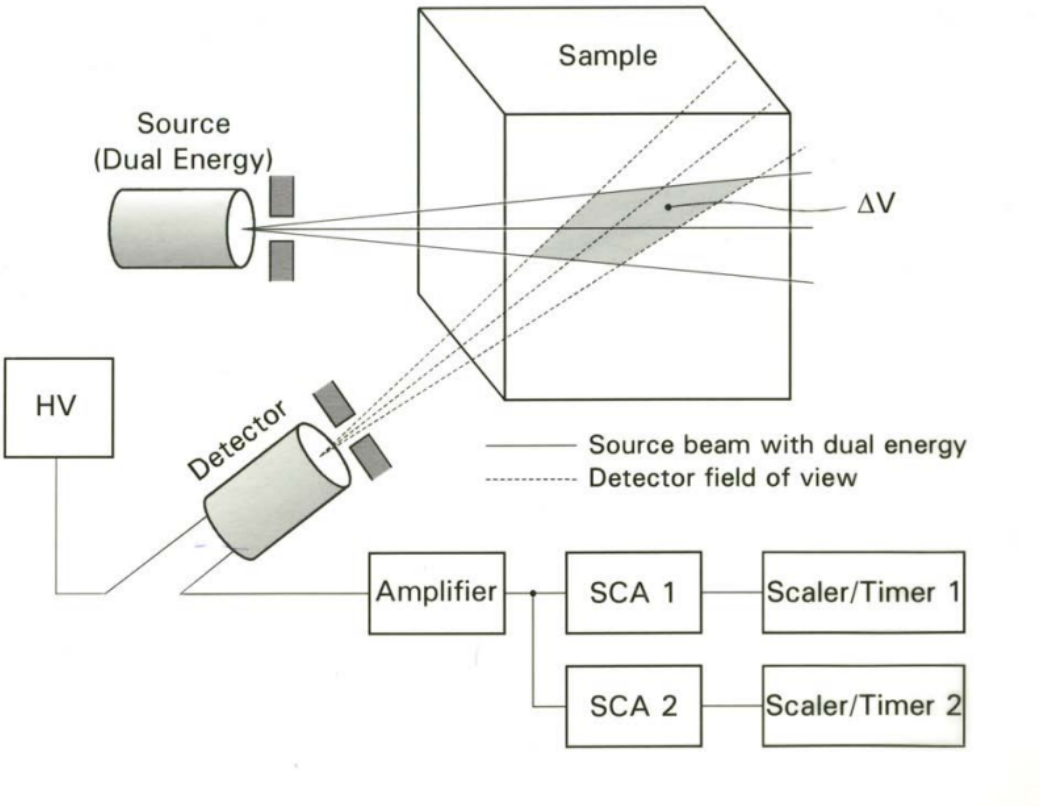


FIGURE 1. Block diagram for the Agricultural Compton scattering tomography. It is used two radioactive source i.e., one operating at 60keV and the second one at 662keV. A high voltage power supply (HV) operating at 960 Volts feeds a scintillator detector (NaI(Tl)). The output signal of the detector is amplified and discriminated by means of two single channel analyzer (SCAs), each one enabled to operate in the region of a given energy, as well as to allow the counters to count the scattered photons under a specific time.

As shown in the above figure, this paper uses cone beam projections. The mathematical problem defined in the paper and the meaning of each term is as follows

$$S(E) = I_0(E)t\varepsilon E)t\left(-\int_{x1}\mu_1(x,E)dx\right)\frac{d_e\sigma^{KN}(E)}{d\Omega}\rho\frac{N_A Z}{A}\exp\left(-\int_{x2}\mu_2(x,E)dx\right)dVd\Omega$$

where,

t - time in seconds during a counting period;

$I_0(E)$ - incident photon flux with energy E;

$\frac{d_e\sigma^{KN}(E)}{d\Omega}$ - Klein-Nishina differential cross-section at energy E for a free electron;

$d\Omega$ - differential angle related to the Klein-Nishina differential cross-section;

ρ - bulk density;

Z - atomic number;

N_A - Avogadro's number;

N - mass number of the material under investigation.

μ_1 and μ_2 - linear attenuation factors of the attenuation of the primary and scattered photons within the sample;

ε - detector's photopeak counting efficiency at the scattered photon energy;

x_1 and x_2 - length of the paths of the photons in the sample from the source to the scattering center and back to the detector respectively, and

dV - differential volume considered for the radiation and its interaction with the matter.

$$\frac{d_e\sigma^{KN}(E)}{d\Omega} = \frac{r_0^2}{2}\left[\frac{1+\cos^2\theta}{(1+\alpha(1-\cos\theta os^2}\left[1+\frac{\alpha^2(1-\cos\theta o^2}{(1+\cos^2\theta)(1+\alpha(1-\cos\theta os}\right]\right]$$

where,

r_0 - classical electron radius numerically equal to 2.818×10^{-15} m;

θ - scattering angle of the scattered photon, and

α - rate between the energy E of the incident photon and the rest mass energy of the electron, which is equal to 511 keV.

(c) The paper uses adaptive filter and 2D wavelets for three dimensional and two dimensional reconstruction algorithms respectively. Additionally, Hamming, wavelets 1D, and the signal restoration were implemented in the module of two dimensional reconstruction.

5. Let $R_\theta(f)$ be the Radon transform of the image $f(x, y)$ in the direction given by θ. Derive a formula for the Radon transform of the scaled image $f(ax, ay)$ where $a \neq 0$ is a scalar. [10 points]

Sol: We know that the Radon transform of the image f(x,y) in direction θ is given by

$$R_\theta(f) = g(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(xcos(\theta) + ysin(\theta) - \rho)dxdy$$

By definition, the formula of the radon transform of the scaled image will be

$$R_\theta^a(f) = g^a(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(ax,ay)\delta(xcos(\theta) + ysin(\theta) - \rho)dxdy$$

Subsituting ax = u, ay =v

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u,v)\delta(\frac{u}{a}cos(\theta) + \frac{v}{a}sin(\theta) - \rho)\frac{du}{a}\frac{dv}{a}$$

$$= \frac{1}{a^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u,v)|a|\delta(ucos(\theta) + vsin(\theta) - a\rho)dudv$$

$$= \frac{|a|}{a^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u,v)\delta(ucos(\theta) + vsin(\theta) - a\rho)dudv$$

$$\implies R_\theta^a(f) = g^a(\rho,\theta) = \frac{g(a\rho,\theta)}{|a|}$$

So the Radon transform of the scaled imaged is nothing but the scaled version (by the same factor) of the radon transform of the original image divided by the square of the factor.

6. Derive the Radon transform of the unit impulse $\delta(x, y)$ and the shifted unit impulse $\delta(x - x_0, y - y_0)$. [10 points]
Sol: We know that the Radon transform of the image f(x,y) in direction θ is given by

$$R_\theta(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(xcos(\theta) + ysin(\theta) - \rho)dxdy$$

a) f(x,y) = δ(x,y)

$$R_\theta(\delta(x,y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x,y)\delta(xcos(\theta) + ysin(\theta) - \rho)dxdy$$

Using the sampling property of delta function,

$$R_\theta(\delta(x,y)) = \delta(xcos(\theta) + ysin(\theta) - \rho)|_{x=0,y=0}$$

$$R_\theta(\delta(x,y)) = \delta(-\rho) = \delta(\rho)$$

Radon transform is always a delta function $\delta(\rho)$ irrespective of θ

b) f(x,y) = δ(x - x_0,y - y_0)

$$R_\theta(\delta(x - x_0, y - y_0)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x - x_0, y - y_0)\delta(xcos(\theta) + ysin(\theta) - \rho)dxdy$$

Using the sampling property of delta function,

$$R_\theta(\delta(x,y)) = \delta(xcos(\theta) + ysin(\theta) - \rho)|_{x=x_0,y=y_0}$$

$$R_\theta(\delta(x,y)) = \delta(x_0cos(\theta) + y_0sin(\theta) - \rho) = \delta(\rho - x_0cos(\theta) - y_0sin(\theta))$$

Radon transform is a delta function that depends on θ

Remarks

The Github folder for this assignment is available here.