Assignment 3: CS 754, Advanced Image Processing

Mohit Kumar Meena-213070021
Shashwat Pathak-213070010

March 22, 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]

Solution:

(a) Root mean squared error (RMSE) after reconstruction: **0.0117**

Figure 1: Original image (left), Noisy image (middle) and reconstructed image (right)

(b) Root mean squared error (RMSE) after reconstruction: **0.0619**

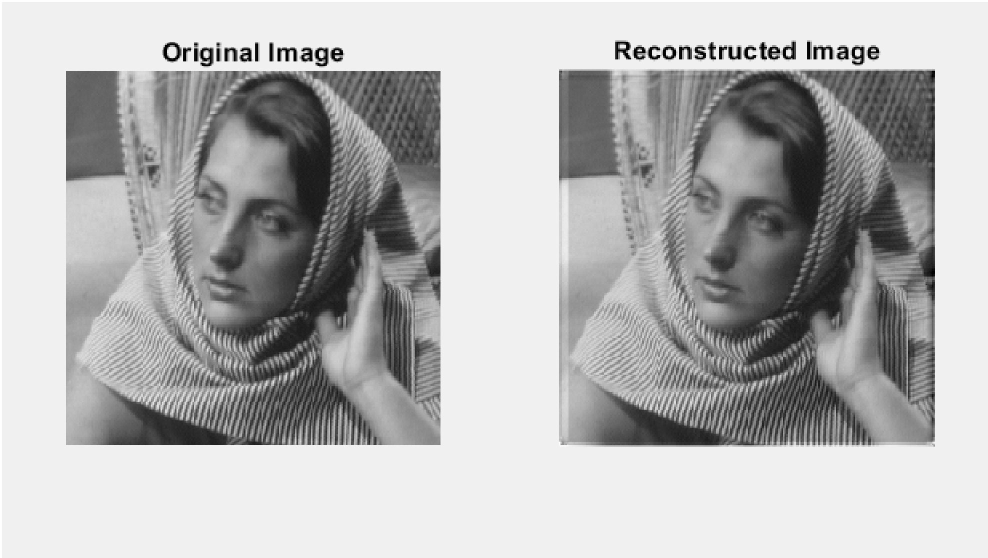


Figure 2: Original image (left) and reconstructed image (right)

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: $J(\boldsymbol{\beta}) = \dfrac{1}{2N}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda_N\|\boldsymbol{\beta}\|_1$ 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 $\lambda_N \geq 2\dfrac{\|\boldsymbol{X}^T\boldsymbol{w}\|_\infty}{N}$ 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]

*****Solutions are pasted below*****

Question-2

Mohit Kumar Meena (213070021)
Shashwat Pathak (213070010)

(a) loss function is defined as

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

and it's always desired that loss function should be strictly convex. Consider a function as given in eqn(1) which is strictly convex w.r.t to set C such that $C \subset R^p$, for all non-zero $\vartheta \in C$.

$$\frac{\vartheta^T \Delta^2 J(\beta) \vartheta}{\|\vartheta\|_2^2} \geq \gamma \quad \text{--- (I)}$$

Here $\Delta^2 J(\beta) = \frac{X^T X}{N}$, $X \in R^{P \times N}$, so eqn (I) can be written as

$$\frac{\vartheta^T X^T X \vartheta}{N \|\vartheta\|_2^2} \geq \gamma \quad \text{for non-zero } \vartheta \in C. \quad \text{--- (II)}$$

As can be seen clearly, the minimum value on LHS is greater than or equal to γ, considering LHS as λ_{min}, we get

$$\lambda_{min} \geq \gamma \quad \text{for all non-zero } \vartheta \in C. \quad \text{--- (III)}$$

Here λ_{min} is the minimum eigenvalue of the corresponding eigenvector in the subset C. So the condition highlighted in eqn (III) is termed as restricted eigenvalue condition.

(b) In the Equation 11.20 (Book), the Lagrangian lasso expression is given as

$$G(u) = \frac{1}{2N} \|y - X(\beta^* + \hat{\vartheta}\|_2^2 + \lambda_N \|\beta^* + \hat{\vartheta}\|_1$$

Also it is given that $\hat{\vartheta} = \hat{\beta} - \beta^*$ is the minimizer of the Eqn (1), so using this assumption, we can conclude that

$$G(\hat{\vartheta}) \leq G(0)$$

(c) To prove:

$$\frac{\|X\hat{\vartheta}\|_2^2}{2N} \leq \frac{w^T X \hat{\vartheta}}{N} + \lambda_N \{ \|\beta^*\|_1 - \|\beta^* + \hat{\vartheta}\|_1 \}$$

Proof:

Earlier, we have proved that $G(\hat{\vartheta}) \leq G(0)$ i.e

$$\frac{\|y - X(\beta^* + \hat{\vartheta})\|_2^2}{2N} + \lambda_N \|\beta^* + \hat{\vartheta}\|_1 \leq \frac{\|y - X\beta^*\|_2^2}{2N} + \lambda_N \|\beta^*\|_1$$

We know that, $y = X\beta^* + w$, so substituting it, we get

$$\frac{\|w - X\hat{\vartheta}\|_2^2}{2N} + \lambda_N \|\beta^* + \hat{\vartheta}\|_1 \leq \frac{\|w\|_2^2}{2N} + \lambda_N \|\beta^*\|_1$$

$$\therefore \frac{\|w - X\hat{\vartheta}\|_2^2}{2N} - \frac{\|w\|_2^2}{2N} \leq \lambda_N (\|\beta^*\|_1 - \|\beta^* + \hat{\vartheta}\|_1) \quad \text{--- (I)}$$

Simplifying $\|w - X\hat{\vartheta}\|_2^2 = (w - X\hat{\vartheta})^T (w - X\hat{\vartheta})$

$$= w^T w - w^T \hat{\vartheta} X - \hat{\vartheta}^T w^T X + \hat{\vartheta} X^T X \hat{\vartheta} \quad \text{--- (II)}$$

substituting eqn (II) in eqn (I), we get

P.T.O

$$\frac{\hat{\vartheta} X^T X \hat{\vartheta}}{2N} - \frac{w^T \hat{\vartheta} X}{N} \leq \lambda_N \left(\|\beta^*\|_1 - \|\beta^* + \hat{\vartheta}\|_1 \right)$$

solving it further, we get

$$\boxed{\frac{\|X\hat{\vartheta}\|_2^2}{2N} \leq \frac{w^T X \hat{\vartheta}}{N} + \lambda_N \left(\|\beta^*\|_1 - \|\beta^* + \hat{\vartheta}\|_1 \right)} \quad \text{— ③}$$

ⓓ Holder's inequality states that

$$\|fg\|_1 \leq \|f\|_p \|g\|_q$$

where $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$

In Eqn (3) above, $w^T X \hat{\vartheta} = \hat{\vartheta}^T X^T w = \langle \hat{\vartheta}, X^T w \rangle$, since it is a scalar, so.

$$\langle \hat{\vartheta}, X^T w \rangle = \sum_{i=1}^{p} \hat{\vartheta}_i X^T w_i \leq \max\{|y_i|\} \sum_{i=1}^{p} |\hat{\vartheta}_i| \quad (y_i = X^T w)$$

we know that maximum is L-∞ norm and absolute is L-1 norm,

so; $\hat{\vartheta}^T X^T w = \langle \hat{\vartheta}, X^T w \rangle \leq \|X^T w\|_\infty \cdot \|\hat{\vartheta}\|_1$ — ④

substituting the results obtained in ④ to Eqn ③, we get

$$\frac{\|X\hat{\vartheta}\|_2^2}{2N} \leq \frac{\|X^T w\|_\infty \cdot \|\hat{\vartheta}\|_1}{N} + \lambda_N \left(\|\beta^*\|_1 - \|\beta^* + \hat{\vartheta}\|_1 \right) \quad \text{— ⑤}$$

Now since $\beta^*_{S_c} = 0$, $\|\beta^*\|_1 = \|\beta_S^*\|_1$ and

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

substituting these results in Eqn (5), we get

$$\boxed{\frac{\|X\hat{\vartheta}\|_2^2}{2N} \leq \frac{\|X^T w\|_\infty \cdot \|\hat{\vartheta}\|_1}{N} + \lambda_N \left(\|\hat{\vartheta}_S\|_1 - \|\hat{\vartheta}_{S^c}\|_1\right)} \quad \text{—} (6)$$

(e) Assuming $\frac{1}{N}\|X^T w\|_\infty \leq \frac{\lambda_N}{2}$, $\|\hat{\vartheta}\|_1 = \|\hat{\vartheta}_S\|_1 + \|\hat{\vartheta}_{S^c}\|_1$

substituting in eqn (6), we get

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

$$\frac{\|X\hat{\vartheta}\|_2^2}{2N} \leq \frac{3}{2}\lambda_N \|\hat{\vartheta}_S\|_1 - \frac{1}{2}\lambda_N \|\hat{\vartheta}_{S^c}\|_1$$

$$\frac{\|X\hat{\vartheta}\|_2^2}{2N} \leq \frac{3}{2}\lambda_N \cdot \|\hat{\vartheta}_S\|_1$$

Now using cauchy-schwartz inequality, we get

$$\boxed{\frac{\|X\hat{\vartheta}\|_2^2}{2N} \leq \frac{3}{2}\sqrt{K}\,\lambda_N \|\hat{\vartheta}\|_2} \quad \text{—} (7)$$

(f) The bound is given as

$$\|\hat{\beta} - \beta^*\|_2 \leq \frac{3}{\gamma}\sqrt{\frac{K}{N}} \cdot \sqrt{N}\,\lambda_N \quad \text{s.t } \lambda_N \geq \frac{2\|X^T w\|_\infty}{N} > 0$$

According to restricted eigenvalue condition;

$$\frac{\vartheta^T X^T X \vartheta}{N\|\vartheta\|_2^2} \geq \gamma \quad \text{for all non-zero } \vartheta \in C$$

i.e $\frac{\vartheta^T X^T X \vartheta}{N} \geq \gamma \cdot \|\vartheta\|_2^2 \Rightarrow \frac{\|X\hat{\vartheta}\|_2^2}{N} \geq \gamma \cdot \|\vartheta\|_2^2$

so using eqn (7), we get

$$\frac{\gamma \|\hat{\vartheta}\|_2^2}{2} \leq \frac{3}{2} \sqrt{K} \lambda_N \|\hat{\vartheta}\|_2$$

$$\|\hat{\vartheta}\|_2^2 \leq \frac{3}{\gamma} \sqrt{K} \cdot \lambda_N \|\hat{\vartheta}\|_2$$

$$\boxed{\|\hat{\vartheta}\|_2 \leq \frac{3}{\gamma} \sqrt{\frac{K}{N}} \sqrt{N} \cdot \lambda_N}$$

(g) In part (e), we have assumed that

$$\frac{\|X^T w\|_\infty}{N} \leq \frac{\lambda_N}{2}$$

which yields

$$\frac{\|X\hat{\vartheta}\|_2^2}{2N} \leq \frac{3}{2} \lambda_N \|\hat{\vartheta}_S\|_1 - \frac{1}{2} \lambda_N \|\hat{\vartheta}_{S^c}\|_1$$

since $\frac{\|X\hat{\vartheta}\|_2^2}{2N} \geq 0$, so we get

$$0 \leq \frac{3}{2} \lambda_N \|\hat{\vartheta}_S\|_2 - \frac{1}{2} \lambda_N \|\hat{\vartheta}_{S^c}\|_1$$

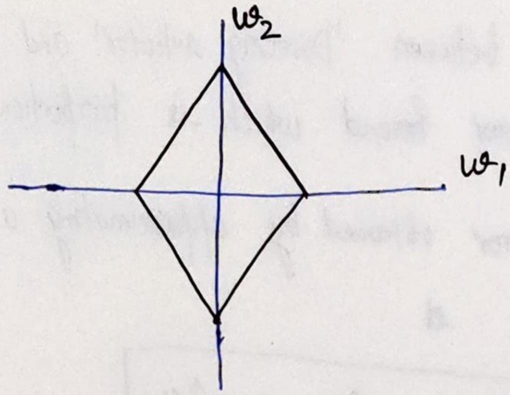
$$\frac{1}{2} \lambda_N \|\hat{\vartheta}_{S^c}\| \leq \frac{3}{2} \lambda_N \|\hat{\vartheta}_S\|_2$$

i.e $\|\hat{\vartheta}_{S^c}\|_1 \leq 3 \|\hat{\vartheta}_S\|_1$

This is termed as cone constraint.

(h) It is always desired that loss function should always be strictly convex but practically this is not always the case, so as we did while defining the restricted eigenvalue condition in part a, that we relaxed a condition for a subset $C \subset \mathbb{R}^p$.

It can be noted that lasso regression always leads to cone constraint i.e $\|\hat{v}_S\|_1 \leq a \|\hat{v}_{S^c}\|_1$, here a is some scalar



(i) Comparision of this particular theorm with theorm 3 is listed below:-

- In this theorm, there is no restriction on parameter (γ) i.e $\gamma > 0$ whereas in theorm 3, $\delta_{2s} < \sqrt{2}-1$.
- In this theorm the possibility of finding minimum eigenvalue is deterministic whereas Theorm 3 relies on explicit calculation of restricted isometry coefficients (RIC).
- Theorm 3 does not account for noise vector correlation with any quantity whereas this particular theorm emphasizes with the noise vector as greater the correlation results greater is the error.
- As described on Page No. 296 "the rate (11.15) - including the logarithmic factor - is known to be minimax optimal,

meaning that it cannot be substantially improved upon by any estimator!!

- But the biggest advantage of theorm 3 is that it handles compressive signals whereas this particular theorm puts the sparsity constraint for the sensed signal which is very tight.

(j) The common thread between 'Dantzig selector' and the LASSO estimation is the error bound which is proportional to $\sqrt{K}$.

As defined, the error obtained by approximating a signal x by some $x \in \Sigma_K$ is

$$\boxed{\sigma_K(x)_p = \min_{x \in \Sigma_K} \|x - \hat{x}\|_p}$$

when $x \in \Sigma_K \Rightarrow \sigma_K(x)_p = 0$ for any value of p.

The error bound on theorm 1.0.1 is obtained to be proportional to $\sqrt{K}$ and by applying same condition on 'Dantzig selector' and using the bounds mentioned, the error bound is proportional to $\sqrt{K}$ as well.

Also the bound in LASSO estimation is proportional to $\lambda_N \geq \frac{2\|X^T w\|_\infty}{N}$, where λ_N is regularization parameter.

Applying similar condition to 'Dantzig selector', error bound is proportional to $\lambda \geq \|A^T e\|_\infty$.

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.

   (a) Filtered-Back Projection

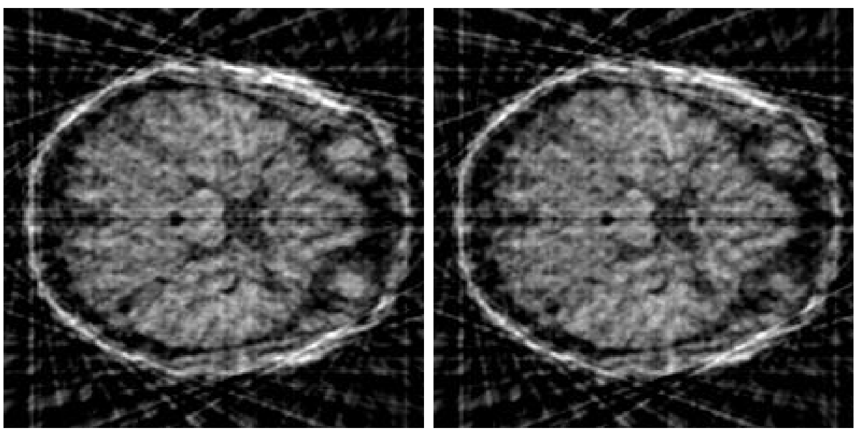


Figure 3: Filtered-Back Projection reconstruction of **Slice 50** and **Slice 51**

   (b) Independent CS based Reconstruction

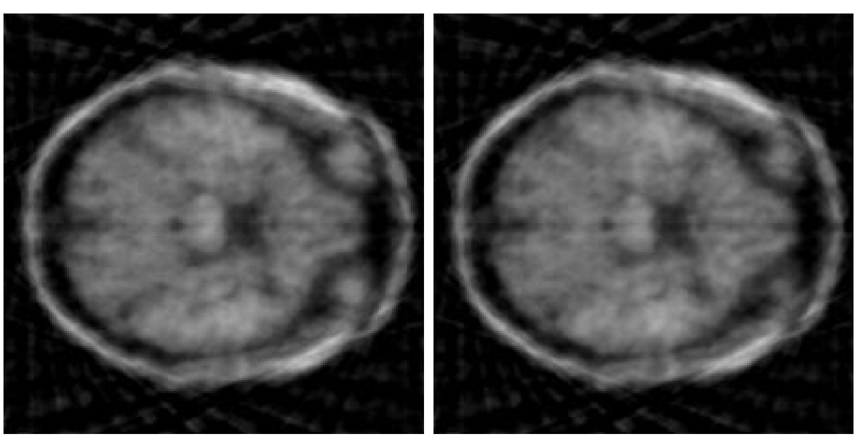


Figure 4: Independent CS based reconstruction of **Slice 50** and **Slice 51**

   (c) Coupled CS Reconstruction using two images (Slices 50 and Slices 51 in this case)

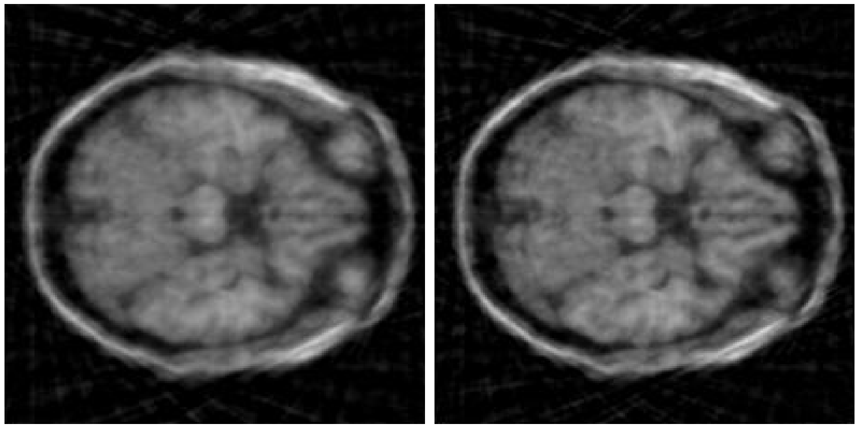


Figure 5: Coupled CS based reconstruction of **Slice 50** and **Slice 51**

(d) Coupled CS Reconstruction using three images(Slices 50, 51 and 52)

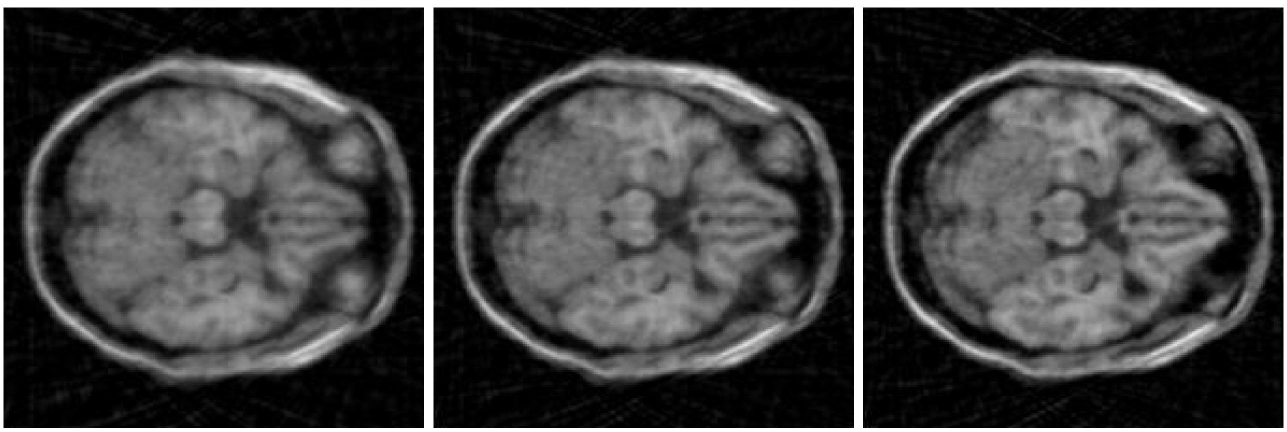


Figure 6: Coupled CS based reconstruction of **Slice 50**,**Slice 51** and **Slice 52**

[3+7+8+7 = 25 points]

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]

Solution

Title : Volumetric additive manufacturing via tomographic reconstruction

Venue : Science Journal

Year of Publication : 2019

Link : `https://www.science.org/doi/abs/10.1126/science.aau7114`

Abstract

This paper implements a manufacturing technique where paralled beam projection from different angles of the desired object are acquired, and back projected on a photosensitive substrate. The interaction of the subtracted with the projected light melts off the material, leaving us with the 3D model of the desired object.

Mathematical Model

The reconstruction was inspired from the back projection technique used in CT reconstruction. The following mathematical equations are described for a z-slice of the desired 3D model $f_T(r, z)$. It takes the values 0 or

1, at every pixel location, depicting the presence or absence of material. The exponential radon transform is represented as follows:

$$f(r,z) = I\frac{\alpha}{\Omega}(T^*_{-\alpha}(g)(r,z)) \geq D_c$$
$$\implies f(r,z) = I\frac{\alpha}{\Omega}(\int_{\theta=0}^{2\pi} g(r.\hat{\theta},\theta,z)\exp -\alpha_T\hat{\theta}_T d\theta) \geq D_c$$

here, I = Indicator function

D_c = threshold

$T^*_{-\alpha}$ = back-projection operator

α = optical absorption coefficient of the resin

Ω = rotation rate of the container

$f(r,z)$ takes values 0 or 1 depicting the presence or absence of material.

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]

Solution

We know that, the formula for computing the radon transform in the continuous signal case is:

$$R_\theta(f(x,y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{1}$$

In equation(1), replacing x, y with ax, ay, we get

$$R_\theta(f(ax,ay)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(ax,ay)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{2}$$

In equation(2), let $x' = ax$ and $y' = ay$, thus we get

$$R_\theta(f(x',y')) = \frac{1}{a^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x',y')\delta(\frac{x'}{a}cos\theta + \frac{y'}{a}sin\theta - \rho)dx'dy' \tag{3}$$

Considering the case when the equation(3) has non zero values. It has non-zero values when $x'cos\theta + y'sin\theta = a\rho$, thus the above equation transforms to :

$$R_\theta(f(x',y')) = \frac{1}{a^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x',y')\delta(x'cos\theta + y'sin\theta - a\rho)dx'dy' \tag{4}$$

Performing change of variables, we get the following result.

$$R_\theta(f(ax,ay)) = \frac{1}{a^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(xcos\theta + ysin\theta - a\rho)dxdy$$

Thus,

$$\boxed{R_{\theta,\rho}(f(ax,ay)) = \frac{1}{a^2}R_{\theta,a\rho}(f(x,y))}$$

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]

Solution

We know that, the formula for computing the radon transform in the continuous signal case is:

$$R_{\theta,\rho}(f(x,y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{5}$$

Replacing $f(x, y) = \delta(x, y)$, we get the following

$$R_{\theta,\rho}(f(x,y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x,y)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{6}$$

Since, $\delta(x, y) = 1$ has a solution only when $x = 0, y = 0$, and the solution is $\delta(\rho)$. The values are 0 otherwise. Thus

$$\boxed{R_{\theta,\rho}(f(x,y)) = \delta(\rho)}$$

In the case of a shifted impulse signal, i.e $\delta(x - x_0, y - y_0)$, upon using the radon transform equation we get

$$R_{\theta,\rho}(\delta(x - x_0, y - y_0)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x - x_0, y - y_0)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{7}$$

Since, $\delta(x - x_0, y - y_0) = 1$ has a solution only when $x = x_0, y = y_0$, and the solution is $\delta(\rho - (x_0cos\theta + y_0sin\theta))$. The values are 0 otherwise. Thus

$$\boxed{R_{\theta,\rho}(\delta(x - x_0, y - y_0)) = \delta(\rho - (x_0cos\theta + y_0sin\theta))}$$