Reconstruction of Similar Images from Poisson Corrupted Sparse Tomographic Projections

-Ankur Mallick
-Roll No: 110110013
-Electrical Engineering (R&D Project)
-Adviser: Prof. Ajit Rajwade

Agenda

- **Motivation**
  - Poisson noise in low intensity CT Images
  - Similarity between successive video frames/slices of human organs
- **Main Result**
- **Mathematical Model**
  - Linearity of Radon Transform
  - Sparsity of natural images in DCT basis and sparsity of frame differences
  - Combining Sparsity and Maximum Likelihood for image reconstruction
- **Initial Guess**
  - Benefits of a good initial guess
  - Dictionary Learning on Radon Transform Image as initial guess
- **Simulation Results**
  - Simulations on video data
  - Simulations on CT slice data
  - Comparison between frame-by-frame reconstruction and joint reconstruction (our results v/s SPIRAL-TAP)

Motivation – CT Imaging

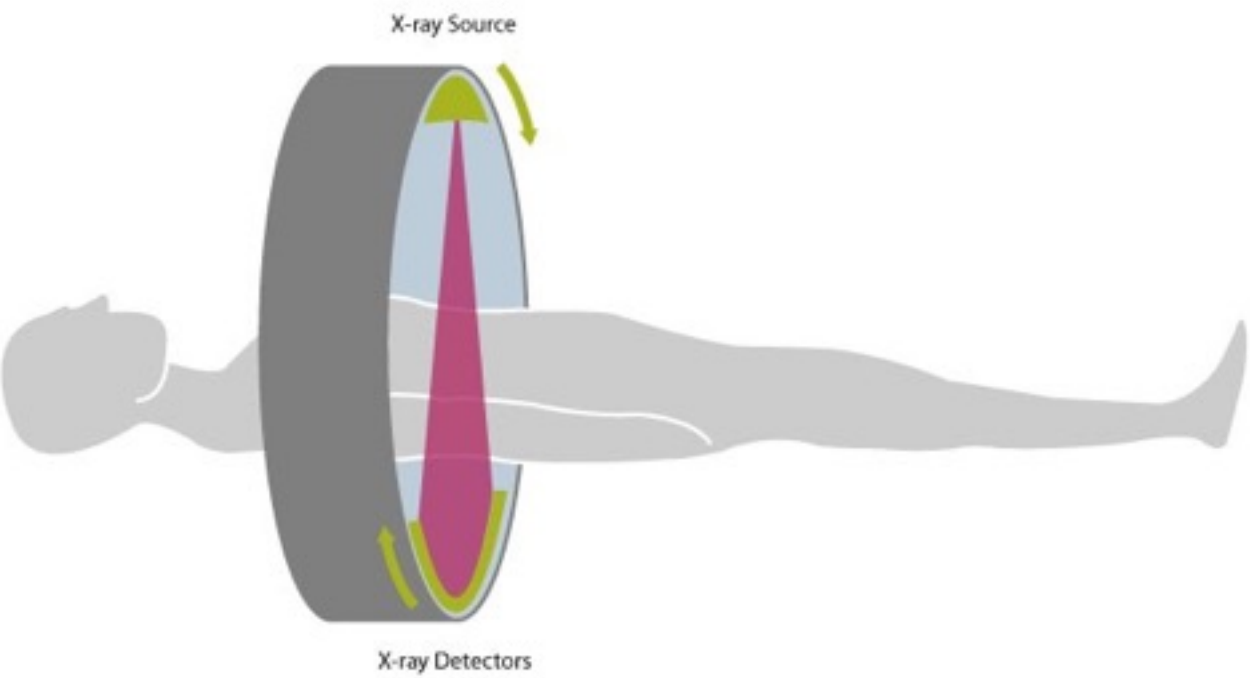
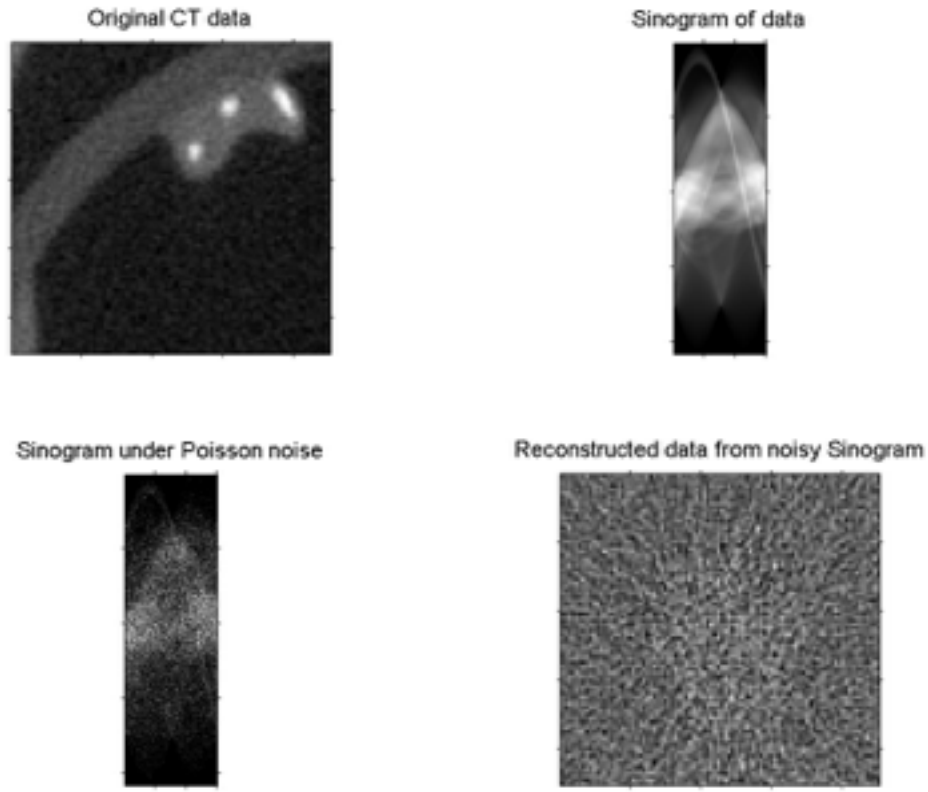


Image Source: http://www.medicalradiation.com/

Motivation – Low Intesity X-Rays

- To reduce the patient's exposure to harmful X-Ray radiation it is desirable to reduce the intensity of X-Rays used for CT Imaging
- At low intensities X-ray detector is highly affected by Poisson noise in the photon counting process
- CT imaging maps the image data to the sinogram (Radon Transform) domain which is known to be corrupted by Poisson Noise for low intensity X-rays
- Received Data =Poisson(Sinogram(Original Data))
- The presence of Poisson noise makes it very difficult to estimate the original image using standard inverse Radon Transform methods such as filtered back projection
- Sparse projections (along only a few angles) are taken to reduce the amount of data

Motivation – Poisson Noise



Effect of Poisson Noise on Filtered Back Projection

Image Source: http://www.sci.utah.edu/cibc-software/ctdata.html

Motivation – Similar Frames/Slices

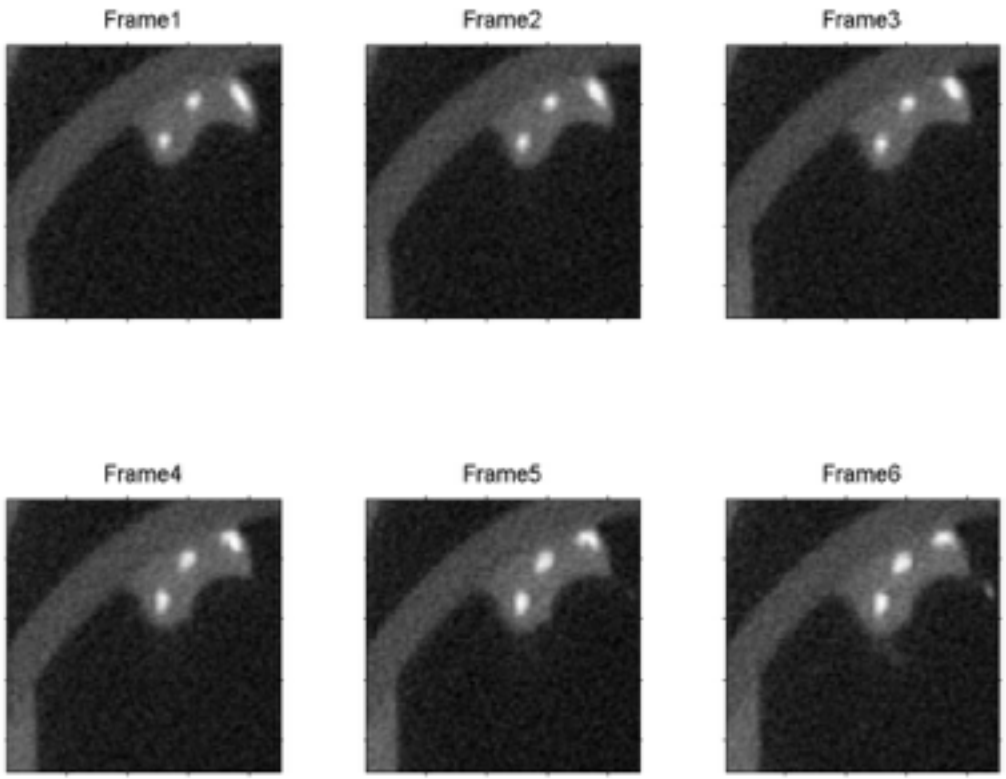
- Most reconstruction algorithms focus on reconstruction of individual images under Poisson noise
- The natural way to apply them to the reconstruction of multiple images (video/organ slices) is to reconstruct each frame individually
- Consecutive frames of a video(with a stable background) or consecutive slices of an organ are extremely similar (differ only in a small fraction of pixels)
- We intend to exploit this similarity in our work to “jointly” reconstruct multiple frames/slices corrupted by organ noise
- Our aim is to achieve a faster and more efficient reconstruction for similar images

Motivation – Similar Frames/Slices

Similarity between consecutive video frames

Image Source: https://media.xiph.org/video/derf/

Motivation – Similar Frames/Slices



Similarity between consecutive organ slices (mouse hand)

Image Source: http://www.sci.utah.edu/cibc-software/ctdata.html

Main Result

- Joint reconstruction of similar video frames/organ slices from their tomographic projections corrupted by Poisson noise by leveraging similarity between successive frames/slices

- An effective initial guess that enables our algorithm to perform better than existing algorithms

- Simulation results showing that the proposed joint reconstruction algorithm performs better than framewise reconstruction and SPIRAL TAP (an existing algorithm for Poisson denoising)

Linearity of Radon Transform

- The radon transform of an image $f(x,y)$ is given by:

$$h(t,\theta)=\int_{(x,y)\in L(t,\theta)} f(x,y)\,dxdy,$$

$$L(t,\theta): x\cos\theta + y\sin\theta = t$$

- Due to linearity of integration we can represent it by the matrix operation:

$$\mathbf{h} = \mathbf{Af}$$

- The matrix A^T defines the unfiltered back projection operation:

$$\mathbf{g} = \mathbf{A}^{\mathrm{T}}\mathbf{h}$$

$$g(x,y) = \int_{\theta=0}^{\pi} h(t(\theta),\theta)\,d\theta$$

$$t(\theta) = x\cos\theta + y\sin\theta$$

Sparsity

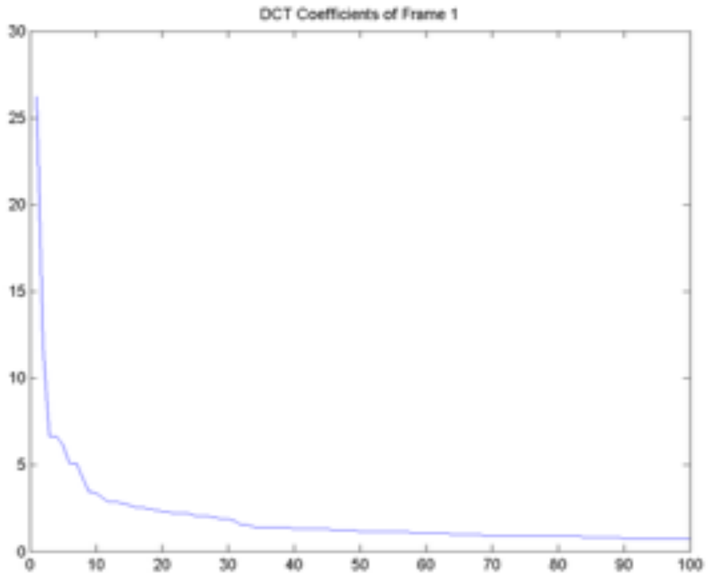
- Natural images are known to be sparse in the DCT basis
- Consider an image vector $\mathbf{f}$ with DCT coefficients vector $\boldsymbol{\theta}$ described by the relation:

$$\mathbf{f} = \mathbf{D}\boldsymbol{\theta}$$

  where $\mathbf{D}$ is the inverse DCT transform matrix (linear operator)

- It is known that $\boldsymbol{\theta}$ is sparse (low L1 norm)
- For two consecutive frames of a video - $\mathbf{f}_1$ (DCT coefficients $\boldsymbol{\theta}_1$) and $\mathbf{f}_2$ (DCT coefficients θ_1) we also expect the vector $|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|$ to be sparse (low L1 norm)

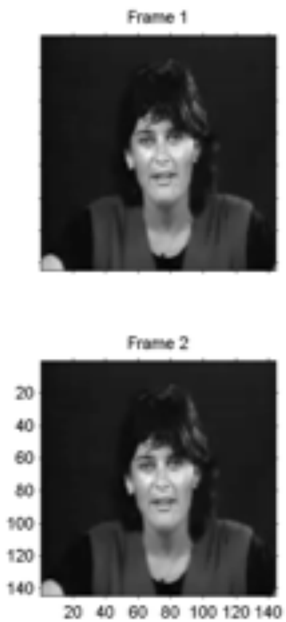
Sparsity



DCT Coefficients of each frame are sparse
(only 1st 100 coeffcients shown, coefficients are plotted in descending order)

Image Source: https://media.xiph.org/video/derf/

Sparsity



DCT Coefficients of the difference between consecutive frames are sparser (only 1st 100 coeffcients shown, coefficients are plotted in descending order)

Image Source: https://media.xiph.org/video/derf/

Mathematical Model

- Consider a video sequence with framewise DCT coefficients $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N$ and framewise tomographic sparse projection matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N$
- Let the Poisson corrupted tomographic projections be $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ which are given by the relation:

$$\hat{\mathbf{y}} = Poisson(\hat{\mathbf{A}}\hat{\boldsymbol{\theta}})$$

$$\hat{\mathbf{y}} = [\mathbf{y}_1 \quad \mathbf{y}_2 \quad . \quad . \quad . \quad \mathbf{y}_N]^{\mathrm{T}}$$

$$\hat{\boldsymbol{\theta}} = [\boldsymbol{\theta}_1 \quad \Delta\boldsymbol{\theta}_1 \quad . \quad . \quad . \quad \Delta\boldsymbol{\theta}_{\mathbf{N}-1}]^{\mathrm{T}}$$

$$\Delta\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k$$

$$\hat{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1\mathbf{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{A}_2\mathbf{D} & \mathbf{A}_2\mathbf{D} & \mathbf{0} & \mathbf{0} \\ . & . & . & . \\ \mathbf{A}_N\mathbf{D} & \mathbf{A}_N\mathbf{D} & \ldots & \mathbf{A}_N\mathbf{D} \end{bmatrix}$$

where $\mathbf{D}$ is the inverse DCT transform matrix

- The Poisson likelihood function is given by:

$$P(\hat{\mathbf{y}} \mid \hat{\mathbf{A}}\hat{\boldsymbol{\theta}}) = \prod_i \frac{(\mathbf{e_i}^T \hat{\mathbf{A}}\hat{\boldsymbol{\theta}})^{\hat{y}_i}}{\hat{\mathbf{y}}_i!} \exp(-\mathbf{e_i}^T \hat{\mathbf{A}}\hat{\boldsymbol{\theta}})$$

where $\mathbf{e}_i$ is the i^{th} canonical basis unit vector

Mathematical Model

- The DCT coefficients are estimated via maximum likelihood with a sparsity prior and a non-negativity constraint on the frame pixel intensities

- The optimization problem for the same is:

$$\min F(\hat{\boldsymbol{\theta}}) + \lambda_1 \| \boldsymbol{\theta}_1 \| 1 + \lambda_2 (\| \boldsymbol{\Delta\theta}_1 \|_1 + \ldots + \| \boldsymbol{\Delta\theta}_{\mathbf{N-1}} \|_1),$$

$$\hat{\mathbf{D}}\hat{\boldsymbol{\theta}} \geq 0$$

$$F(\hat{\boldsymbol{\theta}}) = \mathbf{1}^T \hat{\mathbf{A}}\hat{\boldsymbol{\theta}} - \sum_i \mathbf{y}_i \log(\mathbf{e}_i^T \hat{\mathbf{A}}\hat{\boldsymbol{\theta}} + b)$$

$$\hat{\mathbf{D}} = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{D} & \mathbf{D} & \mathbf{0} & \mathbf{0} \\ . & . & . & . \\ \mathbf{D} & \mathbf{D} & \ldots & \mathbf{D} \end{bmatrix}$$

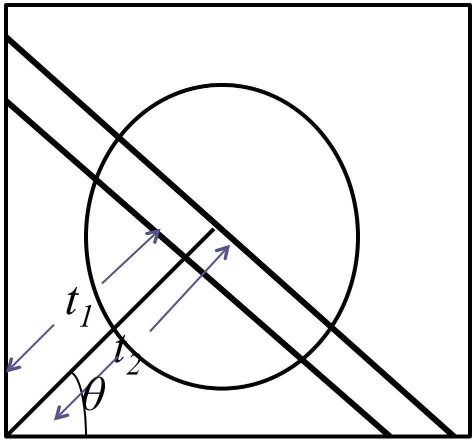
- Here $\mathbf{1}$ is the column vector of all 1's and b is a small constant (10^{-10}) added to ensure the term inside the log does not go to zero
- The regularization parameters (λ_1, λ_2) for the sparsity priors are typically different for $\boldsymbol{\theta}_1$ and $\boldsymbol{\Delta\theta}_\mathbf{k}$ because the DCT coefficients of the latter are sparser

Mathematical Model

- The optimization problem is solved iteratively with each iteration consisting of 3 parts
  1) Gradient Descent:
$$\hat{\boldsymbol{\theta}}_1^{\,l+1} = \hat{\boldsymbol{\theta}}^l - t_l \nabla F(\hat{\boldsymbol{\theta}}^l)$$
$$\nabla F(\hat{\boldsymbol{\theta}}^l) = \hat{\mathbf{A}}^{\mathbf{T}}\mathbf{1} - \sum_i \frac{\mathbf{y}_i}{\mathbf{e}_i^{\,T}\hat{\mathbf{A}}\hat{\boldsymbol{\theta}}^l + b}\hat{\mathbf{A}}^T\mathbf{e}_i$$
  2) Soft thresholding:
$$\hat{\boldsymbol{\theta}}_2^{\,l+1} = \mathrm{sign}(\hat{\boldsymbol{\theta}}_1^{\,l+1}).*[|\hat{\boldsymbol{\theta}}_1^{\,l+1}| - \lambda t_l \mathbf{1}]_+$$
  .* denotes elementwise multiplication and $\lambda = \lambda_1$, λ_2 (for the parts corresponding to $\boldsymbol{\theta}_1$ and $\boldsymbol{\Delta\theta}_\mathbf{k}$ respectively)
  3) Projection onto subspace:
$$\hat{\boldsymbol{\theta}}^{l+1} = \hat{\boldsymbol{\theta}}_2^{\,l} + \hat{\mathbf{D}}^T[-(\hat{\mathbf{D}}^T)^{-1}(\hat{\boldsymbol{\theta}}_2^{\,l})]_+$$
- The step size is changed adaptively to ensure that the objective function decreases at every iteration and the algorithm thus converges

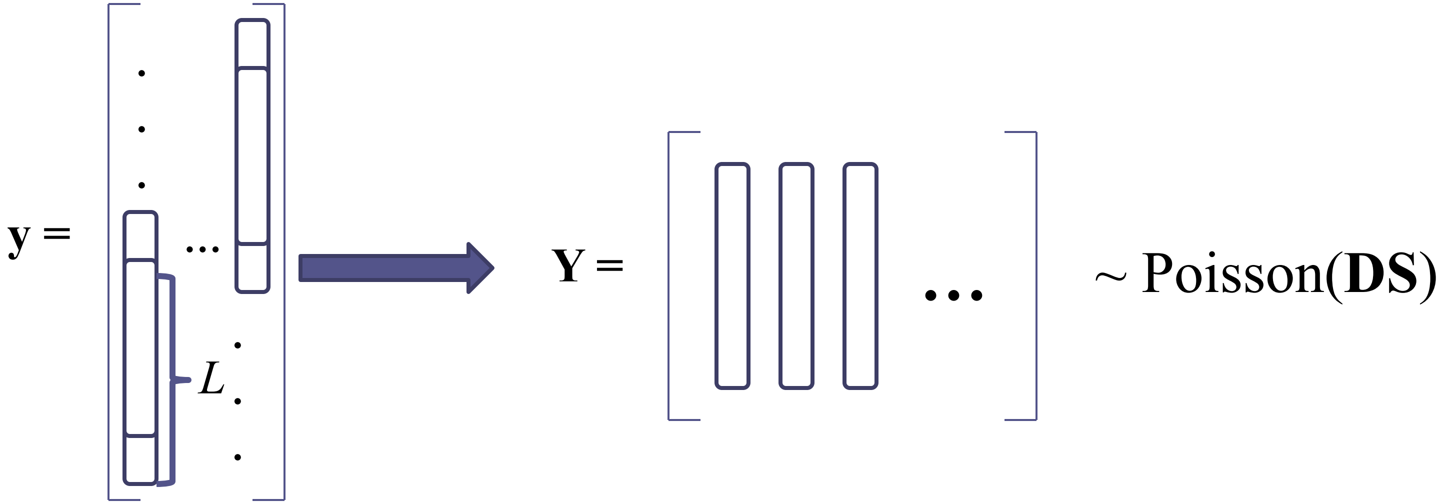
Initial Guess

- A good initial guess (close to the optimal point) is beneficial for any iterative optimization algorithm
- For non-convex functions a good initial guess can help avoid the possibility of getting stuck in a local minimum
- For convex functions it can speed up the convergence to the global minimum
- Consider the figure shown alongside
- Each column of the sinogram of the image corresponds to a particular θ for different values of t
- Thus adjacent elements of any row/column of the sinogram will be highly correlated due to high correlation between neighbouring pixels of the image
- We will be exploiting this correlation in our initial guess



Initial Guess

- Consider the Poisson corrupted sinogram '$\mathbf{y}$' of an image
- Let each column be divided into overlapping vectors of length 'L'
- Let '$\mathbf{Y}$' be the matrix with each column corresponding to one such vector 'patch' from the sinogram
- Due to the similarity between the 'patches' the uncorrupted sinogram is assumed to sparse in some dictionary '$\mathbf{D}$' with sparse coefficients stored in a matrix '$\mathbf{S}$' leading to the following representation:



Initial Guess

- Since the elements of a sinogram are always non-negative we impose the constraints $\mathbf{D} \geq \mathbf{0}$, $\mathbf{S} \geq \mathbf{0}$
- Each column of $\mathbf{D}$ is constrained to have unit L2-norm (standardization)
- The likelihood is given by:

$$P(\mathbf{Y} \mid \mathbf{DS}) = \prod_{i,j} \frac{(\mathbf{DS})_{ij}^{\mathbf{Y}_{ij}}}{\mathbf{Y}_{ij}!} \exp(-(\mathbf{DS})_{ij})$$

- The constrained maximum likelihood optimisation problem can be formulated as:

$$\min \sum_{i,l} \mathbf{Y}_{ij} \log \frac{\mathbf{Y}_{ij}}{(\mathbf{DS})_{ij}} + (\mathbf{DS})_{ij} - (\mathbf{Y})_{ij} + \lambda \sum_{i,l} \mathbf{S}_{ij}$$

$$\mathbf{D} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}, \| \mathbf{D}_k \|_2 = 1$$

- The regularization is due to the sparsity of the coefficients

Initial Guess

- The objective has been slightly modified from the negative log likelihood to give it the form of the KL divergence between $\mathbf{Y}$ and $\mathbf{DS}$
- Hence despite not being jointly convex in $\mathbf{D}$ and $\mathbf{S}$ (individually convex in $\mathbf{D}$ and in $\mathbf{S}$) it can be solved iteratively via alternating minimization
- The update for $\mathbf{S}$ is obtained through the following modified NMF multiplicative update:

$$\mathbf{S}^{k+1} = \mathbf{S}^{k} .* \ (\mathbf{D}^{T}(\mathbf{Y}./(\mathbf{DS}))./(\mathbf{D}^{T}\mathbf{1} \ + \ \lambda)$$

.* and ./ denote elementwise multiplication and division respectively
- The update for $\mathbf{D}$ is obtained through projected gradient descent (step size chosen adaptively):

$$\mathbf{D}_1^{k+1} = [\mathbf{D}^{k} - \alpha(1-(\mathbf{Y}./\mathbf{DS}))\mathbf{S}^{T}]_{+}$$

$$(\mathbf{D}^{k+1})_l = \frac{(\mathbf{D}_1^{k+1})_l}{\|(\mathbf{D}_1^{k+1})_l\|_2}$$

Initial Guess

- The discussed optimization formulation is solved for the sinogram of each frame of the video
- The solution is used to estimate each frame via filtered backprojection
- The estimate is not optimal (because the projections are taken only along a subset of angles in $[0,\pi]$ and hence the correlation between neighbouring patches of the sinogram is not that high)
- However it is good enough for an initial guess to our algorithm
- The initial guess along with the similarity between frames ensures that our algorithm converges much faster than framewise reconstruction

Initial Guess

- SPIRAL TAP, an existing algorithm for reconstructing images from Poisson corrupted projections uses a scaled version of $\mathbf{A}^T\mathbf{y}$ as the initial guess
- Since $\mathbf{A}^T$ corresponds to the unfiltered back projection operation, this gives a very poor and highly blurred estimate and thus affects the performance of the algorithm

$$\mathbf{g} = \mathbf{A}^{\mathrm{T}}\mathbf{h}$$

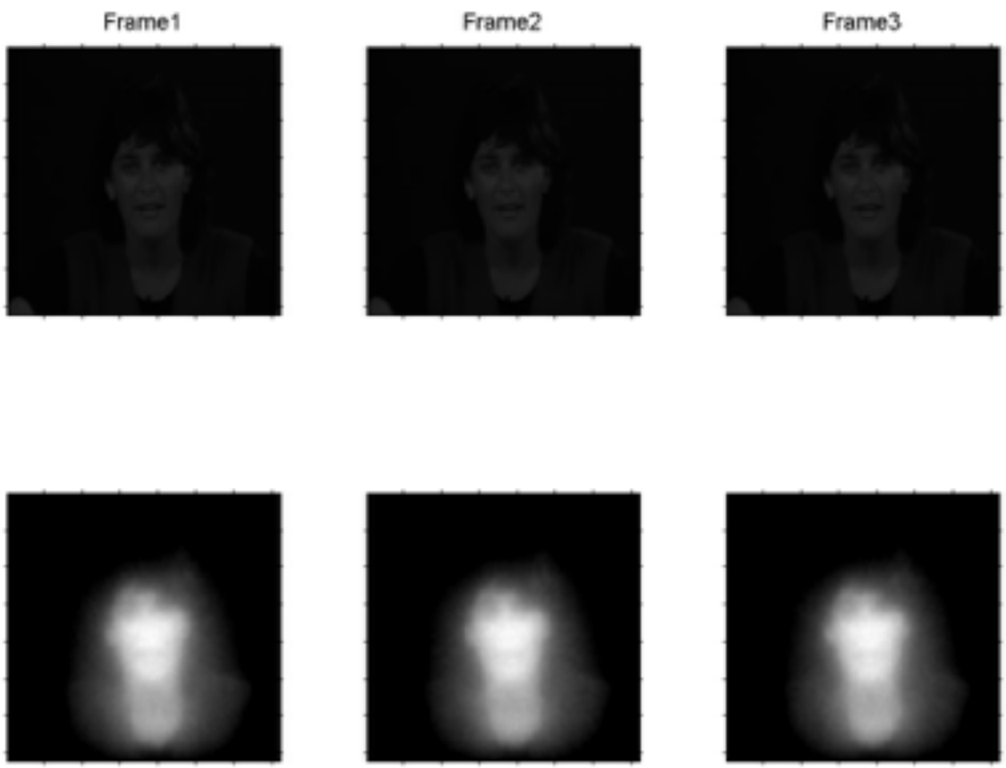
$$g(x, y) = \int_{\theta=0}^{\pi} h(t(\theta), \theta)\, d\theta$$

$$t(\theta) = x\cos\theta + y\sin\theta$$

Simulation Setup

- Two cases – Video sequence ('Miss America'), Organ slices ('Mouse Hand')
- For i^{th} video frame/organ slice, tomographic projections were taken along 60 angles given by the vector (denoting angle in degrees):
  theta = [rem(i,3) : 3 : 179]
- For the initial guess on the corrupted sinogram, the patch length was chosen as $L=10$, and the dictionary $\mathbf{D}$ was assumed to have $5*L$ columns
- The regularisation parameters chosen were:
  $(\lambda_1, \lambda_2) = (10^{-3}, 10^{-4})$
- Comparisons are made between the proposed joint reconstruction algorithm and:
  1) Framewise reconstruction (Same algorithm applied to one frame at a time i.e. $N=1$)
  2) SPIRAL TAP applied framewise (Similar optimisation formulation, different initial guess)
- A total of 6 frames/slices were used but only the first 3 reconstructed frames/slices are shown in the following results

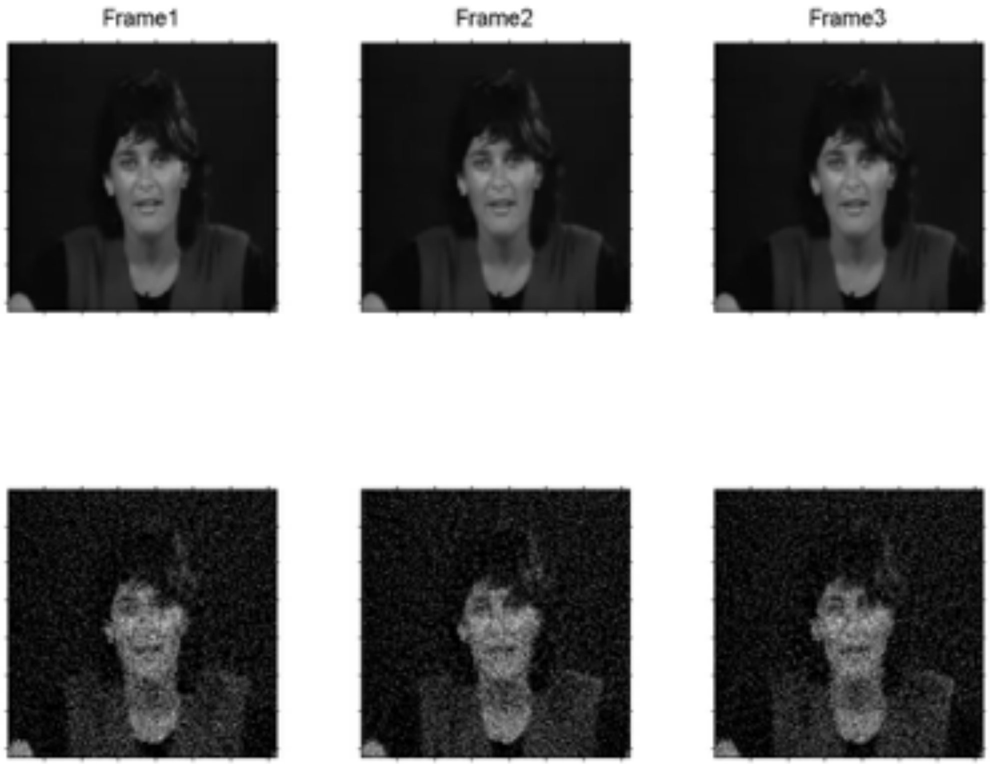
Results - Miss America



SPIRAL TAP Reconstruction of Miss America video sequence

Data for these simulations was procured from https://media.xiph.org/video/derf/

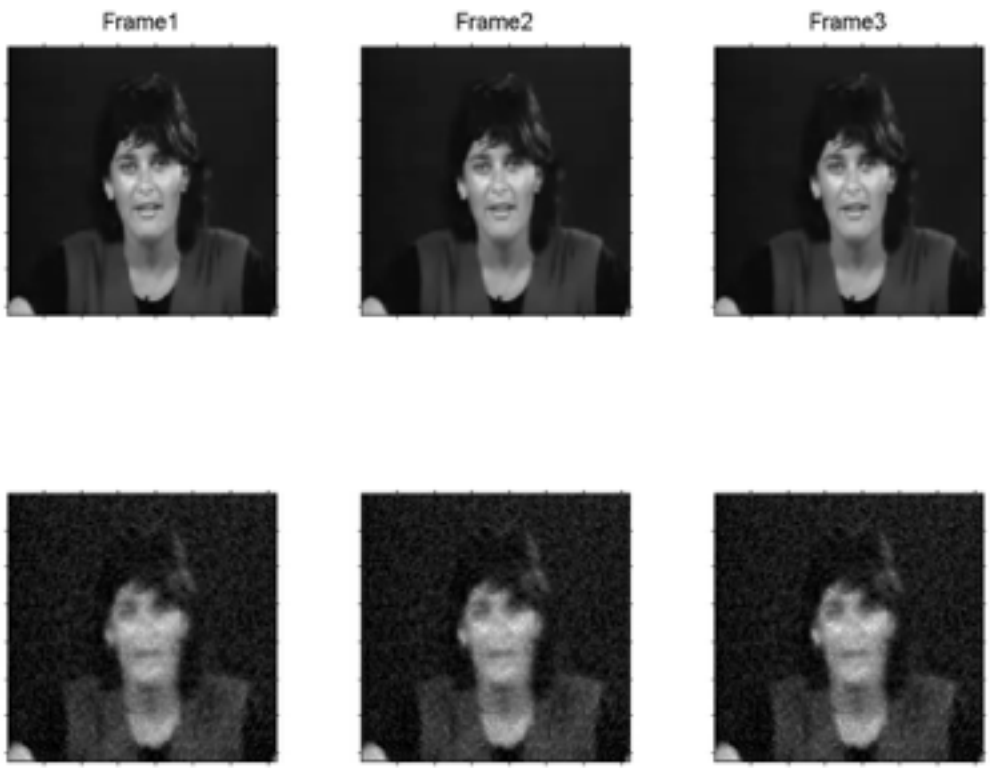
Results - Miss America



Framewise Reconstruction of Miss America video sequence

Data for these simulations was procured from https://media.xiph.org/video/derf/

Results - Miss America



Joint Reconstruction of Miss America video sequence

Data for these simulations was procured from https://media.xiph.org/video/derf/

Results – Miss America

Metric	SPIRAL TAP Reconstruction	Framewise Reconstruction	Joint Reconstruction
PSNR	-0.9514	15.8925	26.9295
RMSE	5.2493	0.7549	0.2119

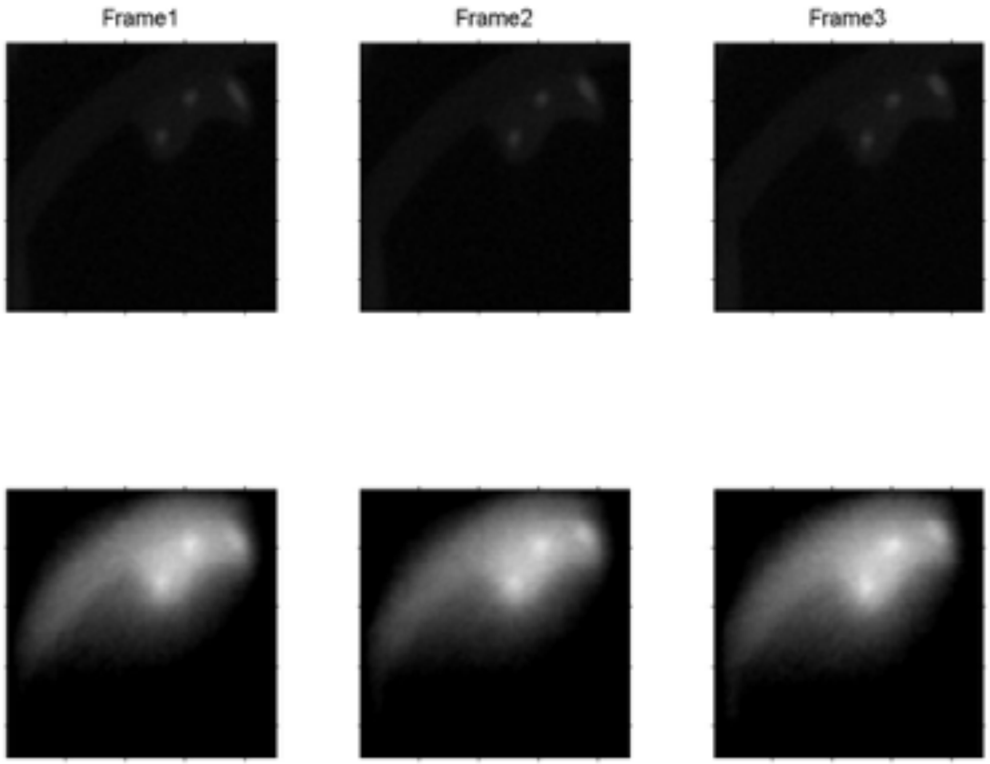
$$E = \| f - f^* \|_2^2$$

$$PSNR = 20 * \log_{10} \max(f^*) - 10 * \log_{10} \frac{E}{N}$$

$$RMSE = \frac{\sqrt{E}}{\| f^* \|_2}$$

f^* = True signal (all frames)
N = Number of pixels in f^*

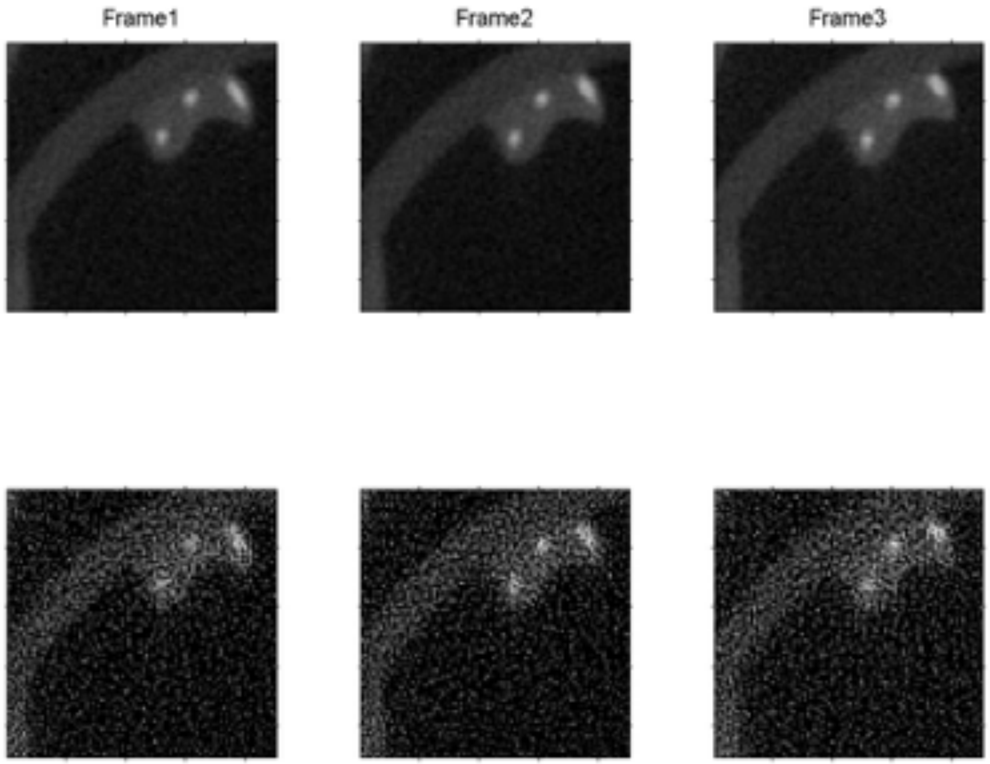
Results – Mouse Hand



SPIRAL TAP Reconstruction of Mouse Hand slices

Data for these simulations was procured from www.sci.utah.edu/cibc-software/ctdata.html

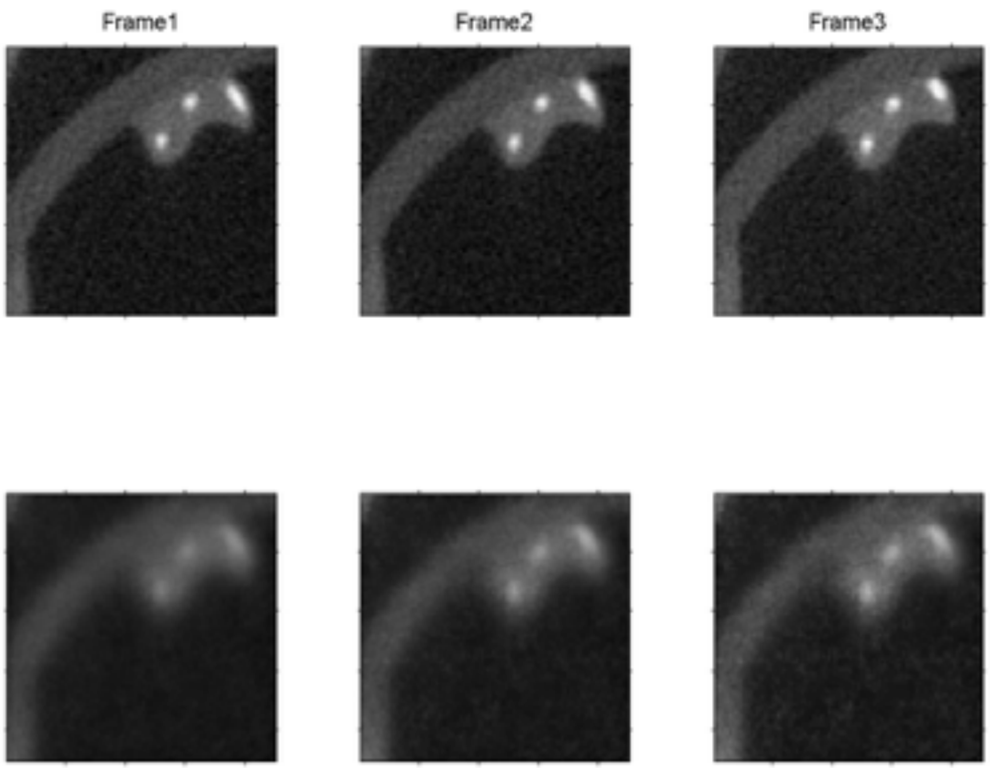
Results – Mouse Hand



Framewise Reconstruction of Mouse Hand slices

Data for these simulations was procured from www.sci.utah.edu/cibc-software/ctdata.html

Results – Mouse Hand



Joint Reconstruction of Mouse Hand slices

Data for these simulations was procured from www.sci.utah.edu/cibc-software/ctdata.html

Results – Mouse Hand

Metric	SPIRAL TAP Reconstruction	Framewise Reconstruction	Joint Reconstruction
PSNR	-3.4125	17.4312	25.3207
RMSE	5.8791	0.5335	0.2151

$$E = \| f - f^* \|_2^{\,2}$$

$$PSNR = 20 * \log_{10} \max(f^*) - 10 * \log_{10} \frac{E}{N}$$

$$RMSE = \frac{\sqrt{E}}{\| f^* \|_2}$$

f^* = True signal (all frames)
N = Number of pixels in f^*

Conclusions

- A joint reconstruction algorithm for Poisson corrupted tomographic projections of video frames or organ slices has been proposed that exploits the high similarity between consecutive frames/slices
- The algorithm performs significantly better than reconstructing each frame/slice individually owing to the larger amount of information obtained due to correlation
- An initial guess involving dictionary learning on the noisy projections has been proposed which enables our algorithm to perform better than an existing method - SPIRAL TAP
- One limitation of the proposed algorithm is that using too many frames might lead to a loss of information about the motion between frames during reconstruction so the number of frames needs to be chosen carefully
- One can possibly extend this work to consider sensing matrices other than the Radon Transform