Compressive Sensing (CS)

Luminita Vese & Ming Yan
lvese@math.ucla.edu yanm@math.ucla.edu

Department of Mathematics
University of California, Los Angeles

The UCLA Advanced Neuroimaging Summer Program (2014)

Background

- Nyquist/Shannon sampling theory: A band-limited signal of interest with highest frequency $\mathbf{B}$ can be exactly reconstructed from its uniformly spaced samples if the rate of sampling exceeds $2\mathbf{B}$ (**Nyquist rate**). This is independently discovered by Kotelnikov, Nyquist, Shannon, and Whitaker.
- The sampling rate needs to be **very high** if the original signal contains high frequencies (to avoid aliasing). The excessive number of samples makes compression necessary prior to storage or transmission. In addition, increasing the sampling rate is very expensive, time consuming (MRI), or dangerous (CT).
- On the other hand, the chance is that most signals we are interested in are highly compressible, namely, they can be represented by a set of sparse or nearly sparse coefficients. CS exploits this feature of signals and thus allows a sampling rate significantly lower than the Nyquist rate.

Efficient image/signal acquisition

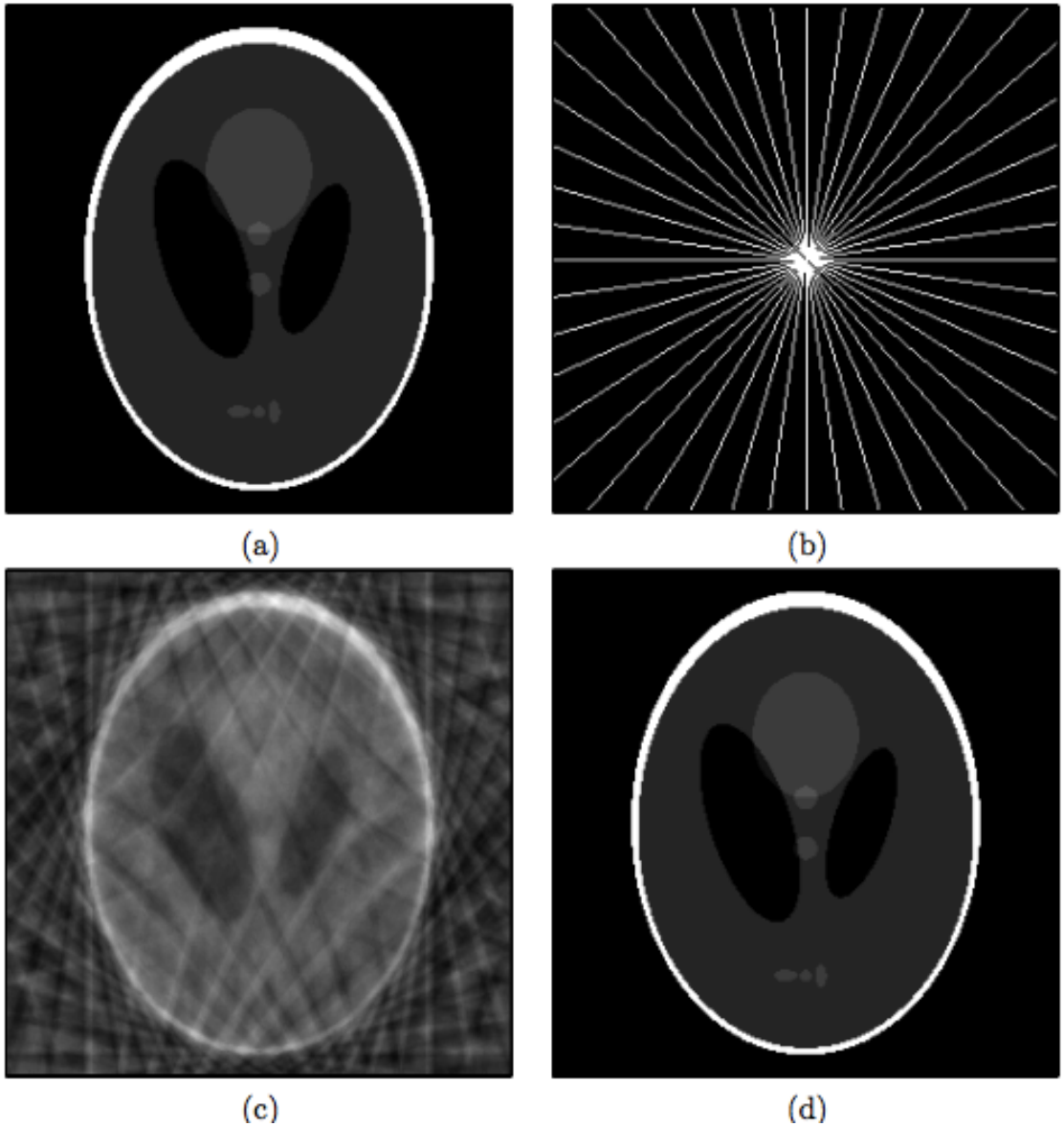
Wish to acquire a digital object $\mathbf{u} \in \mathbf{R}^n$ from m measurements

$$b_k = \langle \mathbf{u}, \mathbf{a}_k \rangle, k = 1, \cdots, m$$

- Few sensors
- Measurements are very expensive
- Measurements process is slow (MRI)
- ...

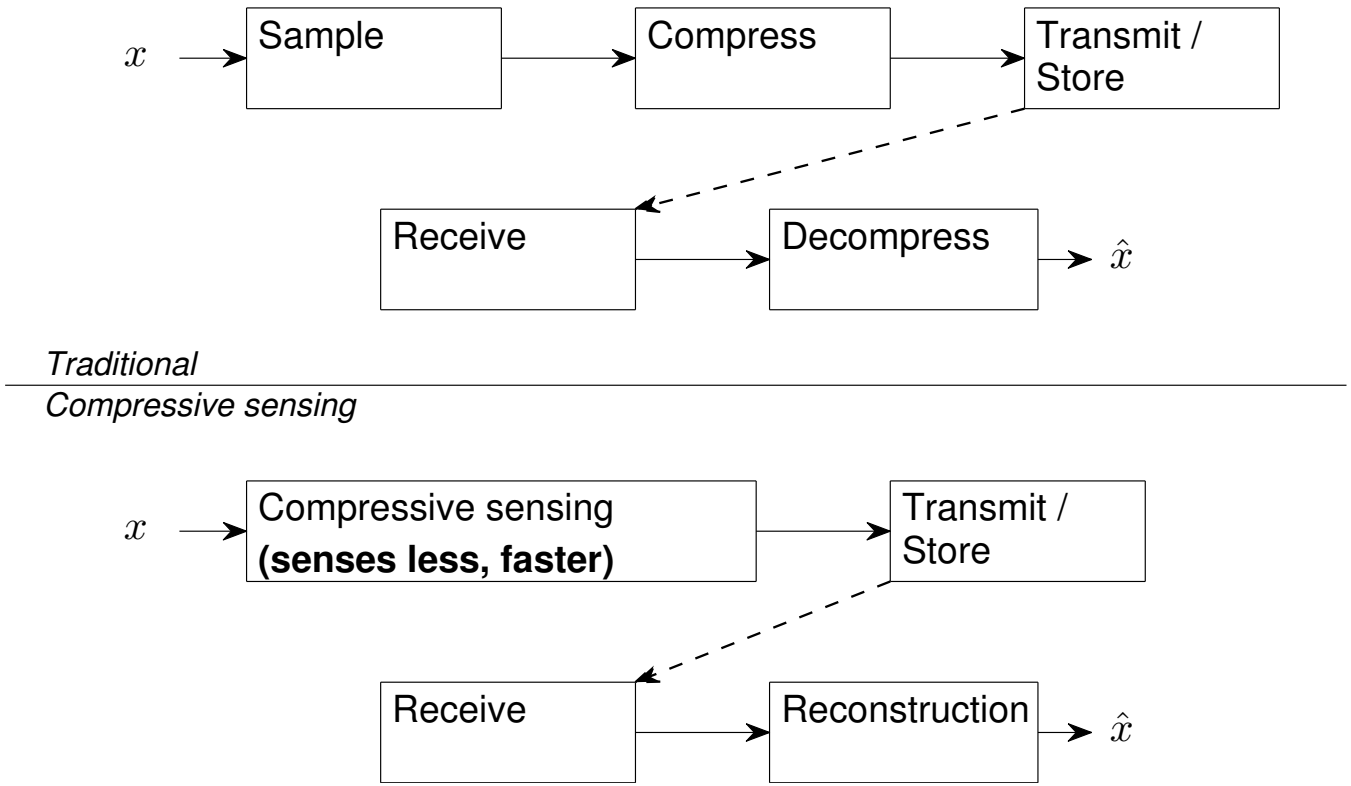
- Is this possible with $m << n$?
- Which measurements should we take?
- How should we reconstruct?

CS-MRI[1]



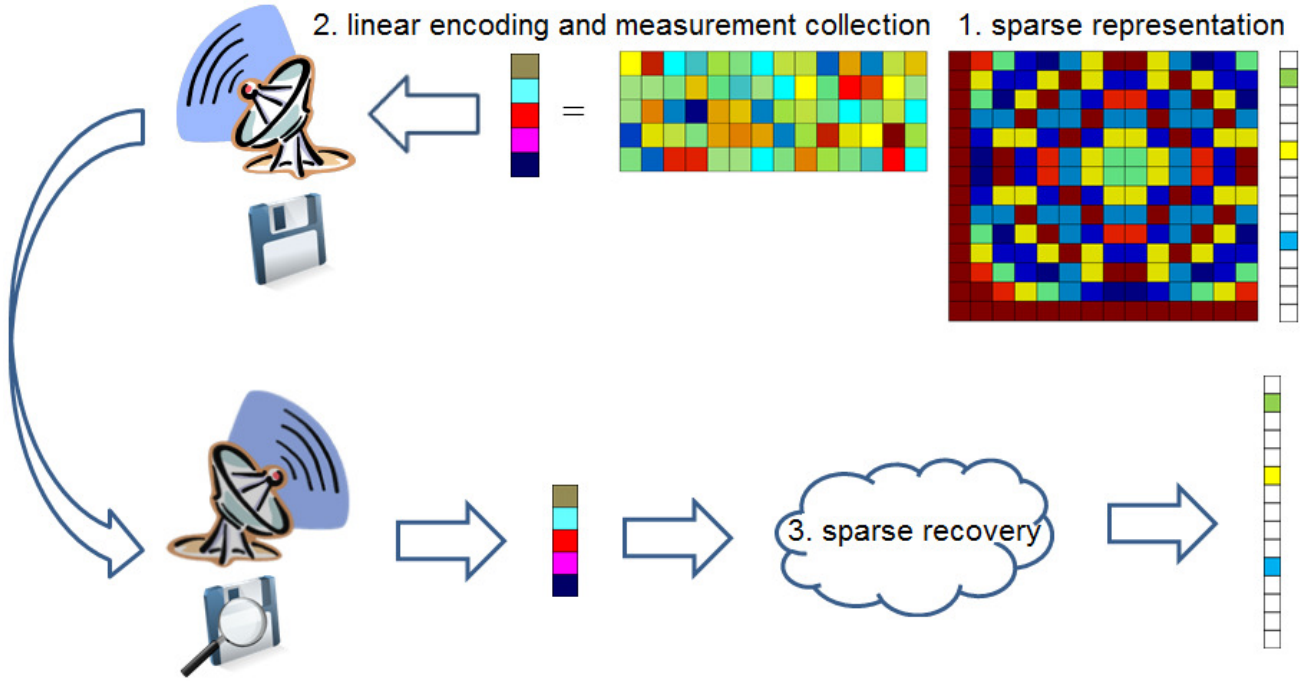
[1] Candes-Romberg-Tao 06'

Traditional sensing versus compressive sensing



The two sensing approaches have their own advantages, address different bottlenecks, fit different needs, and achieve different performances.

Scheme of compressive sensing



- Signal sparse representation
- Linear encoding and measurement collection
- Nonlinear decoding (Sparse recovery)

Sparse representation I

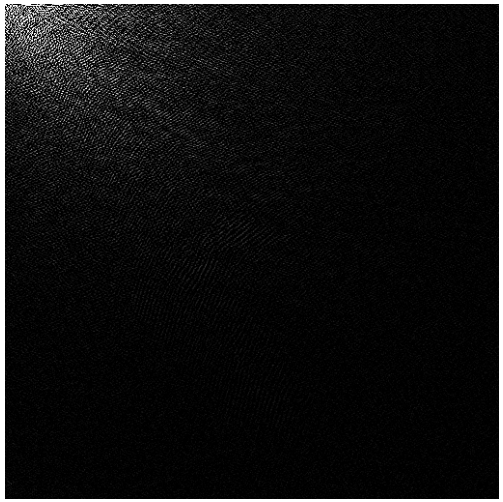
Sparse representation is the basis of CS.

- Express the information of a signal by a small number of real or complex numbers. Mathematically, this is to express a signal $\mathbf{u}^o$ as

$$\mathbf{u}^o = \sum_{i=1}^{p} \psi_i x_i^o,$$

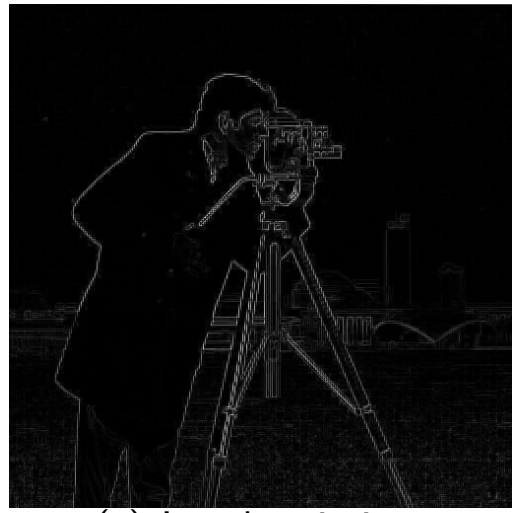
  where all but a small number of entries x_i^o are zero (or small enough to safely neglect). $\Psi = [\psi_1 \ \psi_2 \ \cdots \ \psi_p]$ is called a *dictionary*.
- Besides using a dictionary, a signal can also become sparse under a certain transform Υ, namely, $\Upsilon(\mathbf{u})$ is a sparse vector. Examples include the gradient operator, curvelet transforms, etc.

Sparse representation II

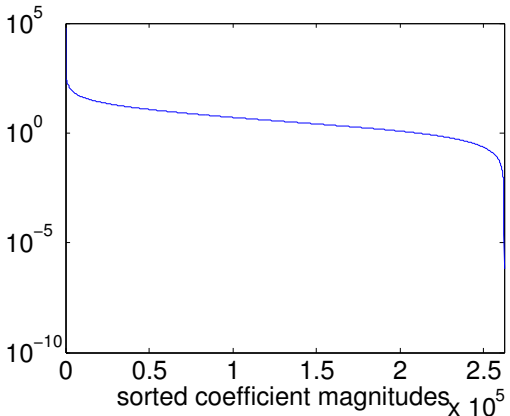


(a) DCT coefficients

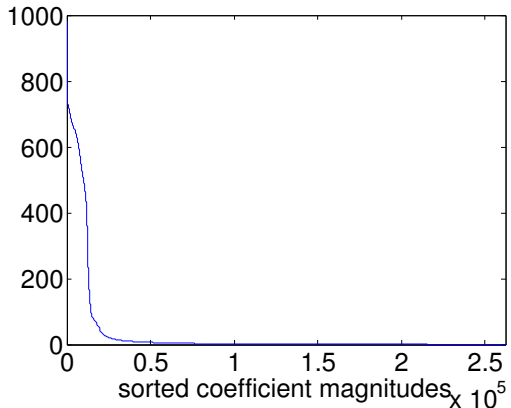
(b) Haar wavelet coefficients



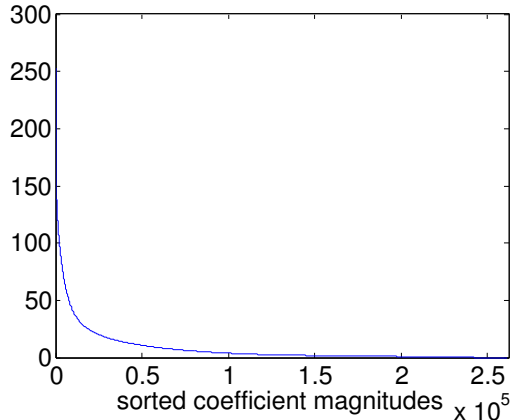
(c) Local variation



(d) DCT coeff's decay



(e) Haar wavelet coeff's decay



(f) Local variation decay

Figure : Sparsity of image Cameraman (the DCT and wavelet coefficients are scaled for better visibility).

CS encoding and decoding

- In CS, the signal $\mathbf{u}^0 = \Psi \mathbf{x}^0$ is encoded to $\mathbf{b} = \mathbf{A}\mathbf{u}^0$. The recovery would be straightforward if $\mathbf{A}$ has full column rank, in which case $\mathbf{u}^0$ would be the unique solution of

$$\underset{\mathbf{u}}{\text{minimize}} \, \|\mathbf{b} - \mathbf{A}\mathbf{u}\|_2^2,$$

  or $\mathbf{u}^0 = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.

- However, CS uses fewer samples and $\mathbf{A}$ has fewer rows than columns. Such a matrix cannot have full column rank, and $\mathbf{b} = \mathbf{A}\mathbf{u}$ has multiple solution.

- What kind of matrix $\mathbf{A}$ allows the recovery of $\mathbf{u}^0$ (or a good approximate of $\mathbf{u}^0$) from $\mathbf{b} = \mathbf{A}\mathbf{u}$ given merely that $\mathbf{x}^0$ is sparse?

Sensing matrix design

Two questions in CS

- How should we design the sensing matrix $\mathbf{A}$ to ensure that it preserves the information in the signal $\mathbf{u}$?
- How can we recover the original signal $\mathbf{u}^0$ from measurements $\mathbf{b}$?

In the case where the data is **sparse** or **compressible**, we can design matrices $\mathbf{A}$ with much fewer rows than columns that ensure the recovery of the original signal accurately and efficiently. This is done in the CS theory of Candes and Tao.

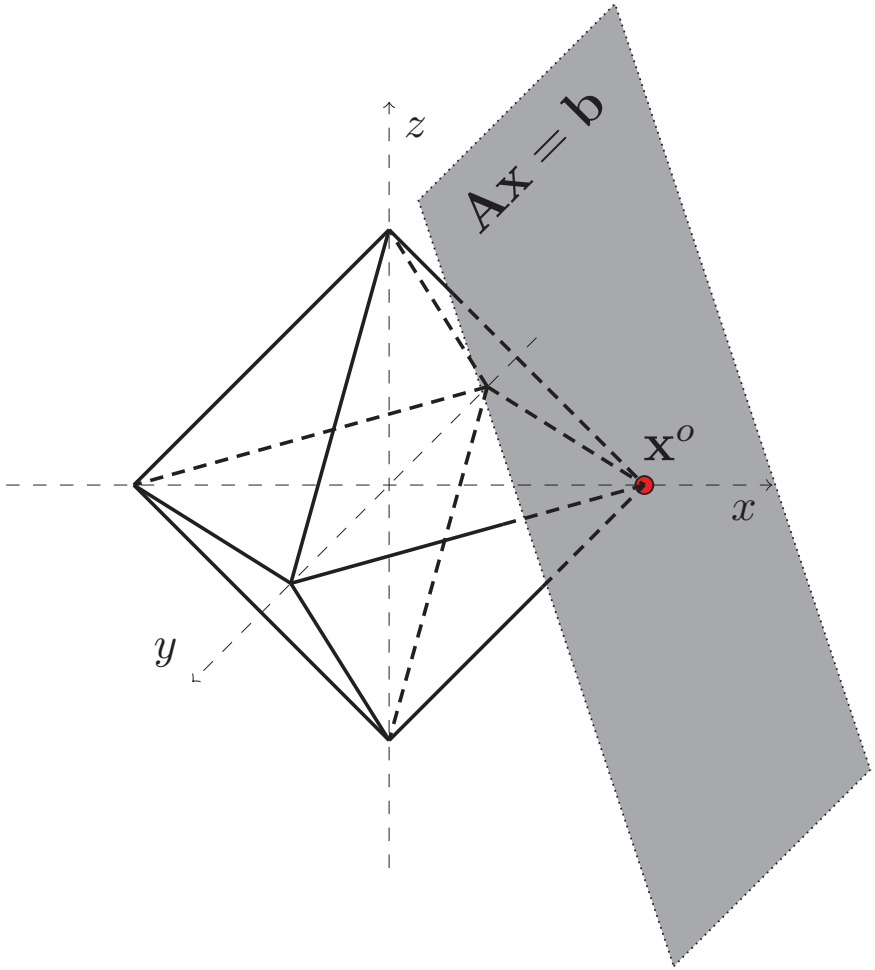
Sensing Matrix

- It is possible to deterministically construct matrices of size $m \times n$ that satisfy the Candes-Tao conditions (RIP condition), but such constructions also require m to be relativey large. [2]
- Fortunately these limitations can be overcome by randomizing the matrix construction.
- It is difficult to verify the conditions.
- We can still use CS even if the conditions are not satisfied: The conditions are sufficient conditions.

[2] DeVore 07', Ubdyk 08'

Basis pursuit

$$\underset{\mathbf{x}}{\text{minimize}}\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = \mathbf{b}\}$$



Basis pursuit denoising and LASSO

$$\underset{\mathbf{x}}{\text{minimize}}\{\|\mathbf{Ax}-\mathbf{b}\|_2 : \|\mathbf{x}\|_1 \leq \tau\}, \tag{1a}$$

$$\underset{\mathbf{x}}{\text{minimize}}\, \|\mathbf{x}\|_1 + \frac{\mu}{2}\|\mathbf{Ax}-\mathbf{b}\|_2^2, \tag{1b}$$

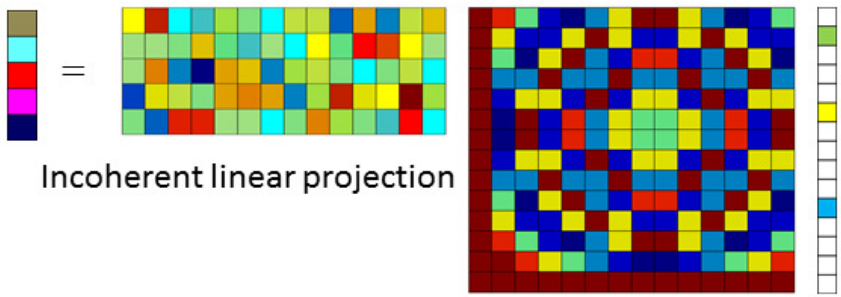
$$\underset{\mathbf{x}}{\text{minimize}}\{\|\mathbf{x}\|_1 : \|\mathbf{Ax}-\mathbf{b}\|_2 \leq \sigma\}. \tag{1c}$$

Questions:

1. Are they equivalent? in what sense?
   - Solution can be non-unique. Why?
   - A solution to one of them is also the solution to the other two with appropriate parameters?
   - Solution sets $\mathcal{X}_\tau = \mathcal{X}_\mu = \mathcal{X}_\sigma$?
2. How to choose parameters?
   - τ, μ, and σ have different meanings.
   - Applications determine which one is easier to set.
   - Use a test data set, then scale parameters for other data.
   - Cross validation

Sparse under basis Ψ

$$\underset{\mathbf{s}}{\text{minimize}}\{\|\mathbf{s}\|_1 : \mathbf{A}\Psi\mathbf{s} = \mathbf{b}\} \tag{2}$$



If Ψ is orthogonal, problem (2) is equivalent to

$$\underset{\mathbf{x}}{\text{minimize}}\{\|\Psi^*\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = \mathbf{b}\}. \tag{3}$$

Also,

$$\underset{\mathbf{x}}{\text{minimize}}\{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 : \|\Psi^*\mathbf{x}\|_1 \leq \tau\},$$

$$\underset{\mathbf{x}}{\text{minimize}}\, \|\Psi^*\mathbf{x}\|_1 + \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2,$$

$$\underset{\mathbf{x}}{\text{minimize}}\{\|\Psi^*\mathbf{x}\|_1 : \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \sigma\}.$$

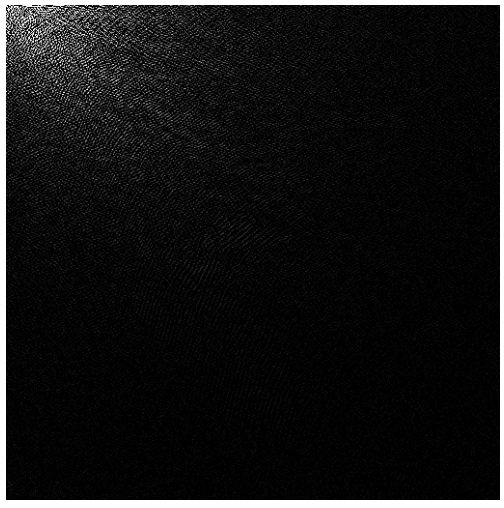
Sparse after transform $\mathcal{L}$

$$\underset{\mathbf{x}}{\text{minimize}}\{\|\mathcal{L}\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = \mathbf{b}\} \tag{4}$$

Examples of $\mathcal{L}$:

- DCT, wavelets, curvelets, ridgelets, ...
- tight frames, Gabor, ...
- (weighted) total variation

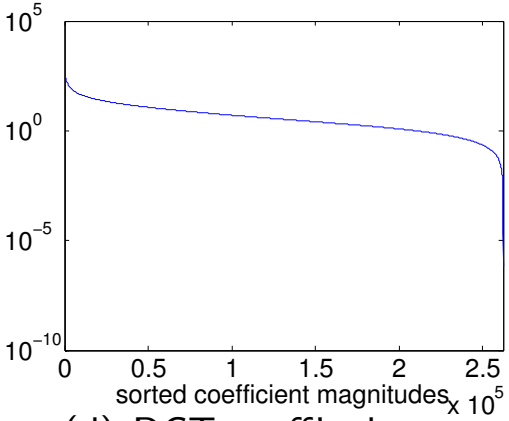
See: E. J. Candès, Y. Eldar, D. Needell and P. Randall. Compressed sensing with coherent and redundant dictionaries. Applied and Computational Harmonic Analysis 31(1), 59–73. ($\mathcal{L}$-RIP $\Rightarrow$ stable recovery of $\mathcal{L}\mathbf{x}$)



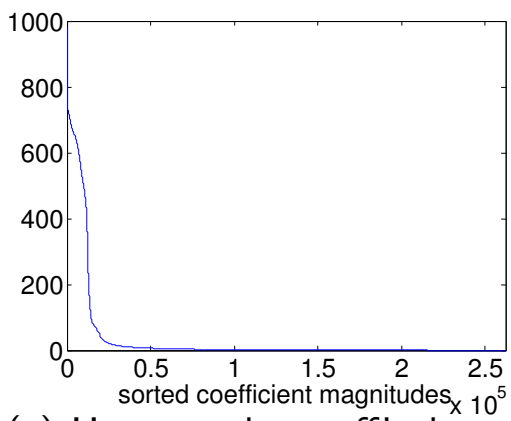
(a) DCT coefficients

(b) Haar wavelet coefficients

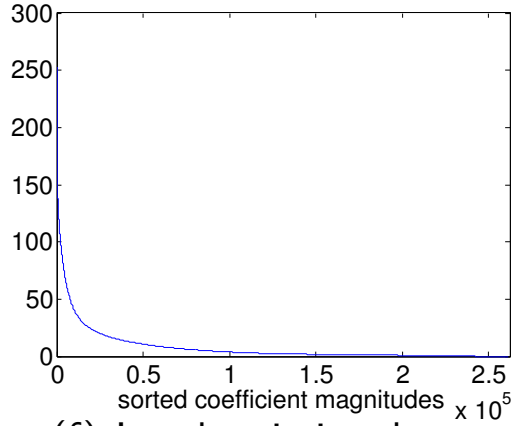
(c) Local variation



(d) DCT coeff's decay



(e) Haar wavelet coeff's decay



(f) Local variation decay

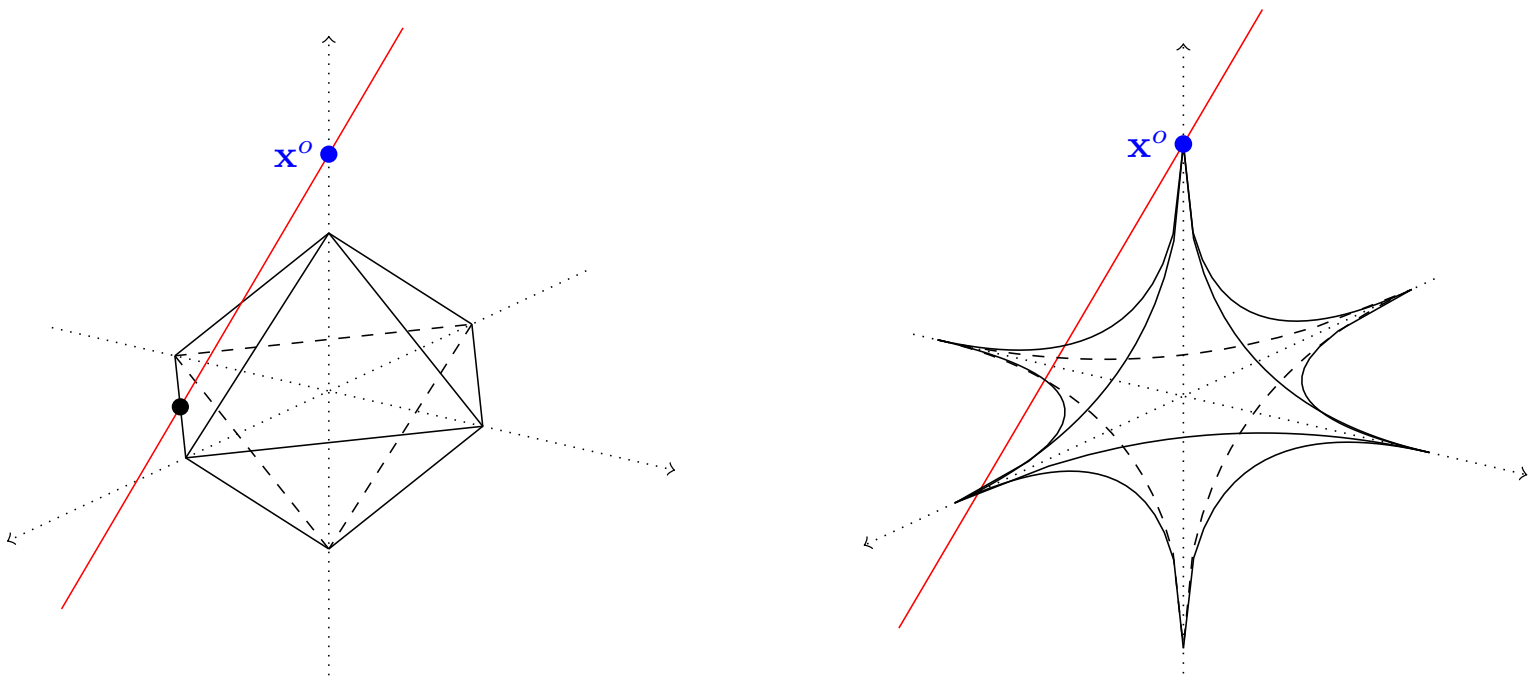
Figure : Sparsity of image Cameraman (the DCT and wavelet coefficients are scaled for better visibility).

Non-convex approaches

- Non-convex optimization, includes ones based on minimizing the non-convex ℓ_q quasi-norm

$$\|\mathbf{x}\|_q = (\sum_i |x_i|^q)^{1/q}, \quad 0 < q < 1,$$

and its variants.



(a) ℓ_1 Minimization (b) $\ell_{1/2}$ Minimization

Figure : ℓ_1 vs. $\ell_{1/2}$ minimization.

Expectation Maximization (EM) & Total Variation (TV) for Computed Tomography (CT): Reconstruction from Undersampled Data

Luminita Vese
Ming Yan

Department of Mathematics, UCLA

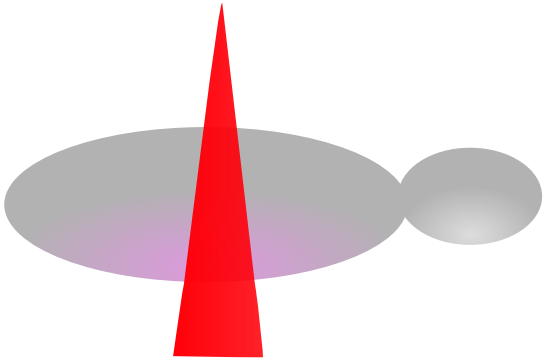
NSF Center for Domain Specific Computing (CDSC), UCLA
Alex Bui and **Jason Cong**

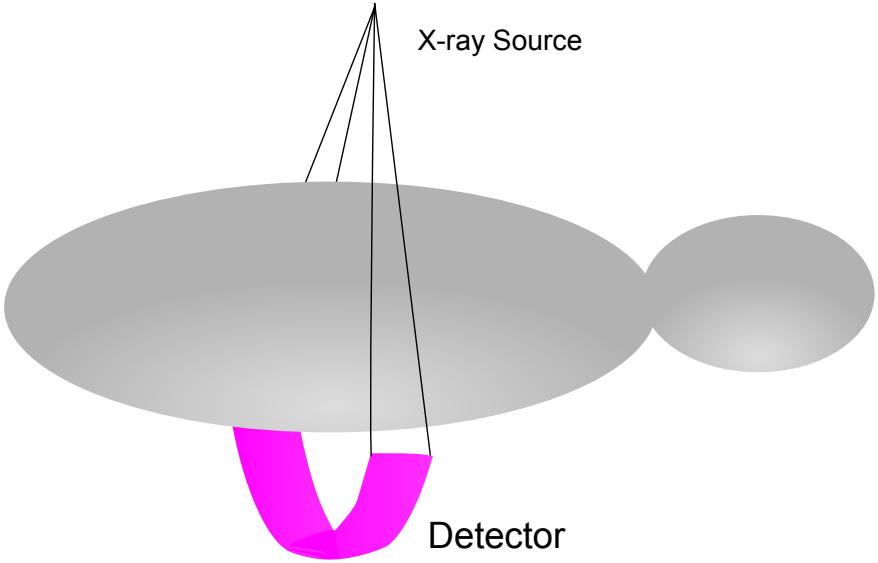
The UCLA Advanced Neuroimaging Summer Program (2014)

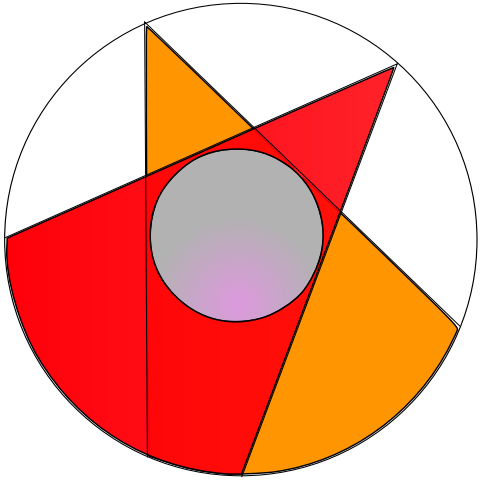
SOMATOM Definition Flash CT System (lowest radiation dose and fastest speed)[1]

[1] www.siemens.com

Cone-Beam Computed Tomography







Attenuation of X-Rays

Notations:

- I_0 is the intensity of the source
- $f(X)$ is the attenuation coefficient of the object at point X
- L is the ray along which radiation propagates
- I is the intensity of radiation at the detector.

Ideal case: These quantities are related by the formula:

$$I = I_0 e^{-\int_L f(X)}.$$

Equivalently

$$\int_L f(X) = \log \frac{I_0}{I}.$$

In practice: Noise, artifacts, and missing data.

Radon Transform

2D: $X = (x, y)$

A straight line L of parameters (θ, t) can be represented as

$$x \cos\theta + y \sin\theta = t.$$

For each pair (θ, t), we have

$$P_\theta(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\delta(x \cos\theta + y \sin\theta - t)dxdy,$$

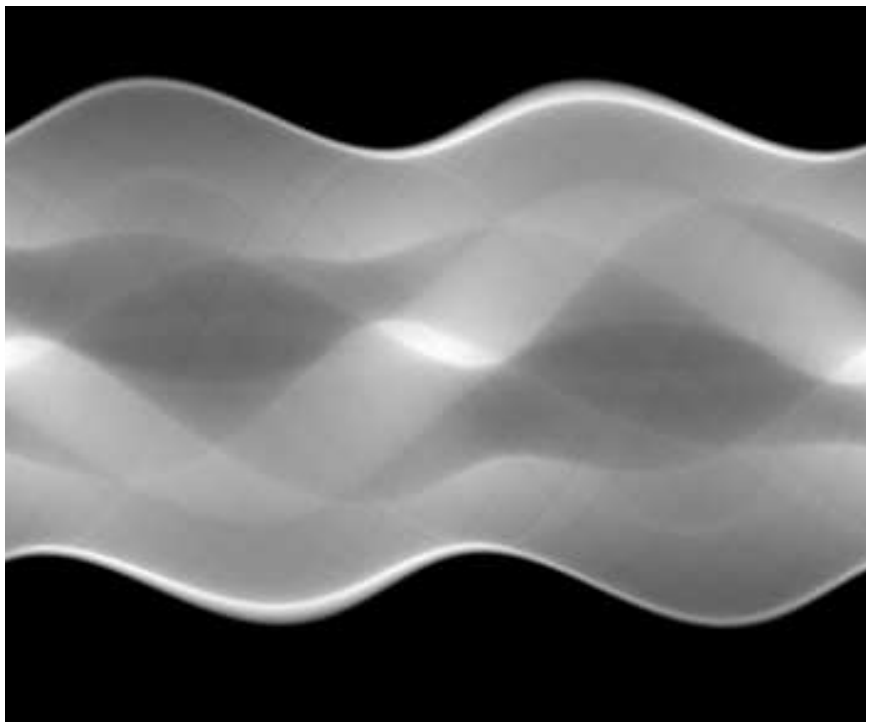
where δ is the Dirac delta function, which is ∞ when (x, y) is on the line $L = \{(x, y) : \ x \cos\theta + y \sin\theta = t\}$, and 0 elsewhere.

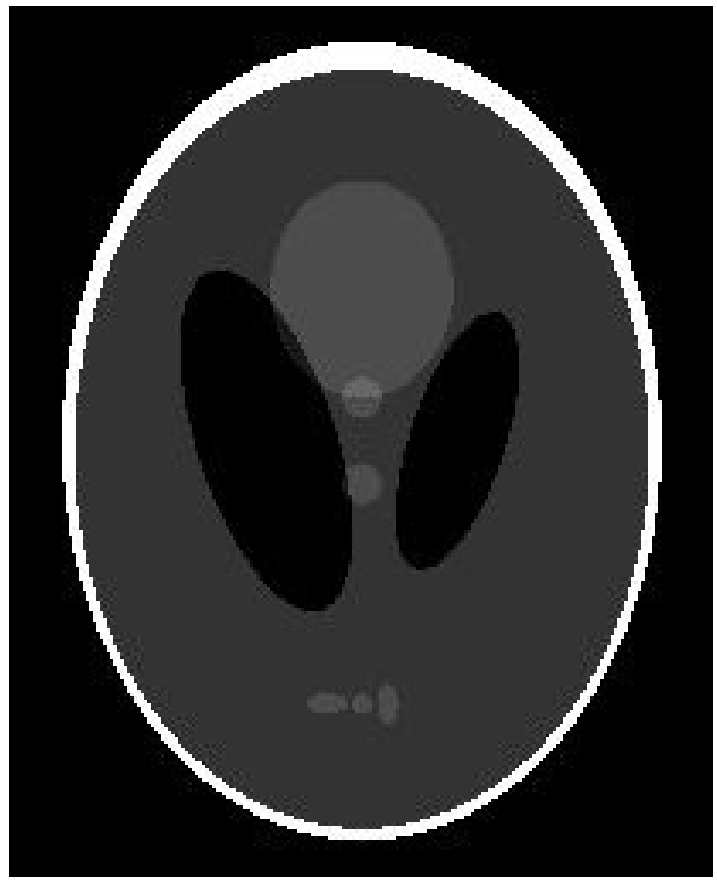
All measurements are stored as sinogram data.

Image Reconstruction

$X = (x, y)$

sinogram data $\left(\int_L f(X) \right)(\theta, t)$ $\Rightarrow$ image $f(X)$





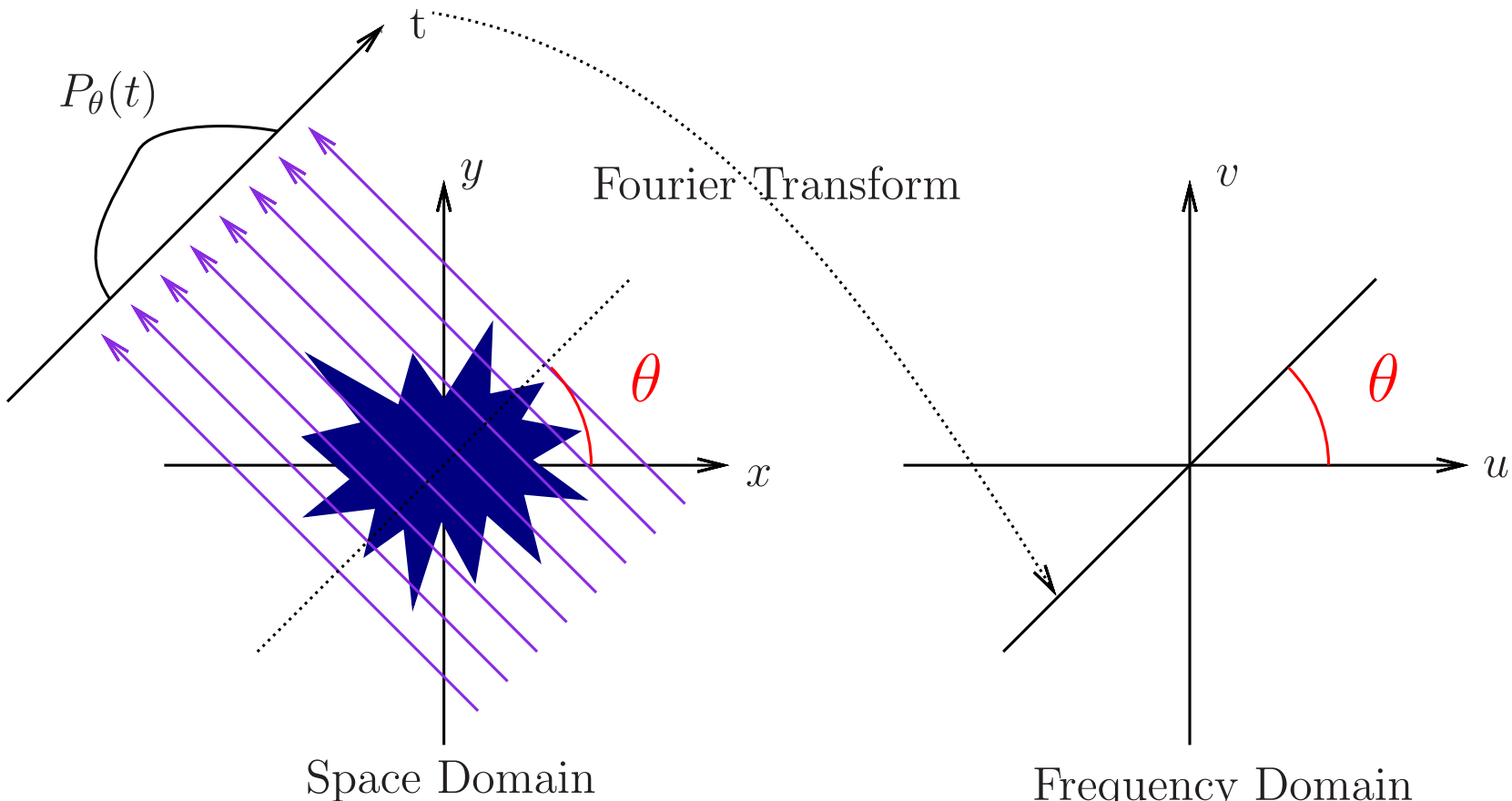
Reconstruction Methods

- Filtered Back Projection (FBP): the most commonly used algorithm in practice by manufacturers. Fast, but it needs many projections.
- Iterative Reconstruction: less sensitive to noise, works when data is incomplete. (Expectation Maximization)
- Compressive Sensing, Total Variation Minimization: Reduced number of views and radiation.

Filtered Back Projection: Fourier Slice Theorem

Using $\mathcal{P}(w,\theta) = F(u,v)|_{u=w\cos\theta, v=w\sin\theta} = F(w\cos\theta, w\sin\theta)$ then

$$f(x,y) = \int_0^{\pi} \left(\int_{-\infty}^{+\infty} |w| \mathcal{P}(w,\theta) e^{2\pi i w t} dw \right)|_{t=x\cos\theta + y\sin\theta} d\theta$$



Iterative Reconstruction

The reconstruction problem is to estimate an image vector x from the following system of equations

$$b = Ax + n,$$

where b is the known measurement vector, A is the discrete Radon transform, and n is the unknown noise in the measurements (note also that the noise n is not always additive).

Iterative Reconstruction is an iterative procedure, producing a sequence of vectors $x^0, x^1, \cdots$ converging to x^*, an approximate solution to the inverse problem.

Expectation Maximization (EM)[2]

Expectation Maximization is the method based on maximizing the probability to observe the given results in the coincidence detectors. For CT, the type of noise can be modeled by a Poisson distribution, thus:

$$P(b|Ax) = \prod_{i=1}^{M} e^{-a_i x} \frac{(a_i x)^{b_i}}{b_i!},$$

where a_i is the i^{th} row of A.
Instead of considering the probability directly, we minimize

$$-\log P(b|Ax) = \sum_{i=1}^{M} (a_i x - b_i \log(a_i x)) + const,$$

with a constraint that $x \geq 0$.

[2]Shepp and Vardi, 1982

Expectation Maximization (cont'd)

For this constraint problem, the Karush-Kuhn-Tucker (KKT) condition is

$$\sum_{i=1}^{M}(a_{ij}(1-\frac{b_i}{a_i x}))-y_j=0, \qquad j=1,\cdots,N,$$

$$y\geq 0, \quad x\geq 0, \quad y^T x=0.$$

Thus, $y_j * x_j = 0$ and

$$x_j\sum_{i=1}^{M}(a_{ij}(1-\frac{b_i}{a_i x}))=0, \qquad j=1,\cdots,N.$$

Therefore, we have the following iterative procedure (EM)

$$x_j^{n+1}=\frac{\sum_{i=1}^{M}(a_{ij}(\frac{b_i}{a_i x^n}))}{\sum_{i=1}^{M}a_{ij}}x_j^n. \tag{1}$$

Total Variation Regularization (TV) [3]

The total variation $\int |\nabla x|$ became very popular by its success in image denoising, and is extensively used in many other practical applications.
The discrete TV regularization proposed by Rudin, Osher, and Fatemi (ROF) for a 2D image x is

$$J(x) = \sum_{i,j} \sqrt{|x_{i+1,j} - x_{i,j}|^2 + |x_{i,j+1} - x_{i,j}|^2},$$

which is isotropic and not differentiable.
To make it differentiable, we can choose

$$J(x) = \sum_{i,j} \sqrt{|x_{i+1,j} - x_{i,j}|^2 + |x_{i,j+1} - x_{i,j}|^2 + \epsilon},$$

where $\epsilon > 0$ is a small number.

[3] Rudin, Osher, Fatemi, 1992

Prior related work

- T. Le, R. Chartrand, T. Asaki, *T. Le, R. Chartrand and T. Asaki, A Variational Approach to Constructing Images Corrupted by Poisson Noise*, 2007
- Christoph Brune et al., *Forward-Backward EM-TV methods for Inverse Problems with Poisson noise Christoph Brune*, 2011
- A. Sawatzky, C. Brune, J. Müller and M. Burger, *Total Variation Processing of Images with Poisson Statistics*, 2009
- X. Zhang et al.
- Work of Wotao Yin, and of Rick Chartrand (nonconvex methods) on compressive sensing

EM+TV

Combining EM and TV, with TV being a penalty term, we have the optimization problem,

$$\begin{aligned} &\underset{x}{\text{minimize}} \ \beta \int |\nabla x| - \log P \\ &\text{subject to } x \geq 0, \end{aligned}$$

with $P = \prod_{i=1}^{M} e^{-a_i x} \frac{(a_i x)^{b_i}}{b_i!}$. Or

$$\begin{aligned} &\underset{x}{\text{minimize}} \ \beta \int |\nabla x| + \sum_{i=1}^{M} (a_i x - b_i \log(a_i x)) \\ &\text{subject to } x \geq 0. \end{aligned}$$

This is a convex problem, and we can use many algorithms for solving it.

EM+TV (cont'd)

The KKT condition of this is

$$-\beta \mathrm{div}(\frac{\nabla x}{|\nabla x|})_j + \sum_{i=1}^{M}(a_{ij}(1 - \frac{b_i}{a_i x})) - y_j = 0, \qquad j = 1, \cdots, N,$$

$$y \geq 0, \quad x \geq 0, \quad y^T x = 0.$$

Similarly, it becomes

$$-\beta \frac{x_j}{\sum_{i=1}^{M} a_{ij}} \mathrm{div}(\frac{\nabla x}{|\nabla x|})_j + \frac{\sum_{i=1}^{M}(a_{ij}(1 - \frac{b_i}{a_i x}))}{\sum_{i=1}^{M} a_{ij}} x_j = 0, \qquad j = 1, \cdots, N.$$

or

$$-\beta \frac{x_j}{\sum_{i=1}^{M} a_{ij}} \mathrm{div}(\frac{\nabla x}{|\nabla x|})_j + x_j - \frac{\sum_{i=1}^{M}(a_{ij}(\frac{b_i}{a_i x}))}{\sum_{i=1}^{M} a_{ij}} x_j = 0, \qquad j = 1, \cdots, N.$$

EM+TV (cont'd)

Denote

$$x_j^{EM} = \frac{\sum_{i=1}^{M}(a_{ij}(\frac{b_i}{a_i x}))}{\sum_{i=1}^{M} a_{ij}} x_j,$$

and the KKT condition becomes

$$-\beta \frac{x_j}{\sum_{i=1}^{M} a_{ij}} \mathrm{div}(\frac{\nabla x}{|\nabla x|})_j + x_j - x_j^{EM} = 0, \qquad j = 1, \cdots, N,$$

which is the optimality condition for the following TV minimization problem

$$\underset{x}{\text{minimize}} \; E_1^p(x, x^{EM}) \equiv \beta \int |\nabla x| + \sum_{j=1}^{N}(\sum_{i=1}^{M} a_{ij})(x_j - x_j^{EM} \log x_j). \quad (2)$$

Let $\sum_i a_{ij} = v_j$ (long vector v_j, or v_{ij} in the image notation)

EM+TV (cont'd)

The EM+TV optimality condition can be solved using an iterative semi-implicit scheme (given the EM step). In 2D we view $x = (x_{i,j})$

$$
\begin{aligned}
&- \beta \frac{x^n_{i,j}}{V_{i,j}} \frac{x^n_{i+1,j} - x^{n+1}_{i,j}}{\sqrt{\epsilon + (x^n_{i+1,j} - x^n_{i,j})^2 + (x^n_{i,j+1} - x^n_{i,j})^2}} \\
&+ \beta \frac{x^n_{i,j}}{V_{i,j}} \frac{x^{n+1}_{i,j} - x^n_{i-1,j}}{\sqrt{\epsilon + (x^n_{i,j} - x^n_{i-1,j})^2 + (x^n_{i-1,j+1} - x^n_{i-1,j})^2}} \\
&- \beta \frac{x^n_{i,j}}{V_{i,j}} \frac{x^n_{i,j+1} - x^{n+1}_{i,j}}{\sqrt{\epsilon + (x^n_{i+1,j} - x^n_{i,j})^2 + (x^n_{i,j+1} - x^n_{i,j})^2}} \\
&+ \beta \frac{x^n_{i,j}}{V_{i,j}} \frac{x^{n+1}_{i,j} - x^n_{i,j-1}}{\sqrt{\epsilon + (x^n_{i+1,j-1} - x^n_{i,j-1})^2 + (x^n_{i,j} - x^n_{i,j-1})^2}} + x^{n+1}_{i,j} - x^{EM}_{i,j} = 0.
\end{aligned}
$$

Algorithm

Input: Given x^0, ϵ, $k = 0$;
while $k < Num\_Iter$ & $\|x^k - x^{k-1}\| < \epsilon$ **do**
  $k = k + 1$;
  $x^{k-\frac{1}{2}} = EM(x^{k-1})$ using (1) ;
  $x^k = \text{argmin } E_1^p(x, x^{k-\frac{1}{2}})$ by solving (2);
end

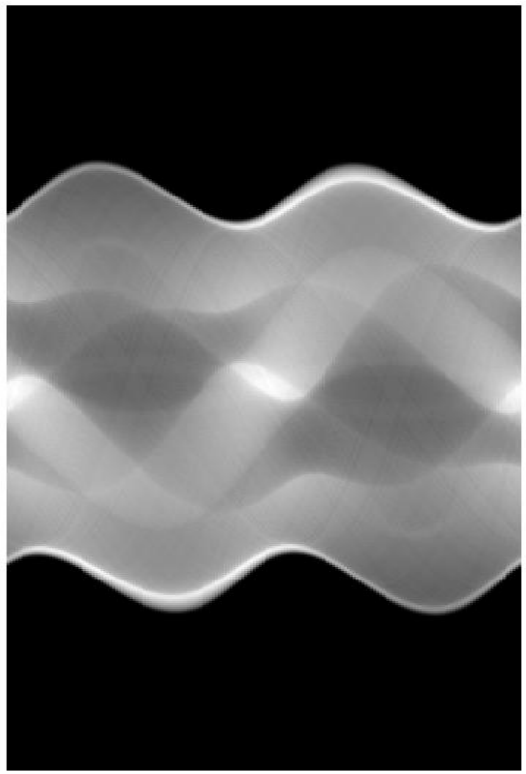
- Convergence Analysis (shown by Ming Yan).
- Positivity: the positivity will remain unchanged in both EM and TV steps.

2D Reconstruction Results

Figure : phantom image

To construct the sinogram data, the fan-beam geometry is used, and for each view, we choose 301 measurements.

Sinogram Data



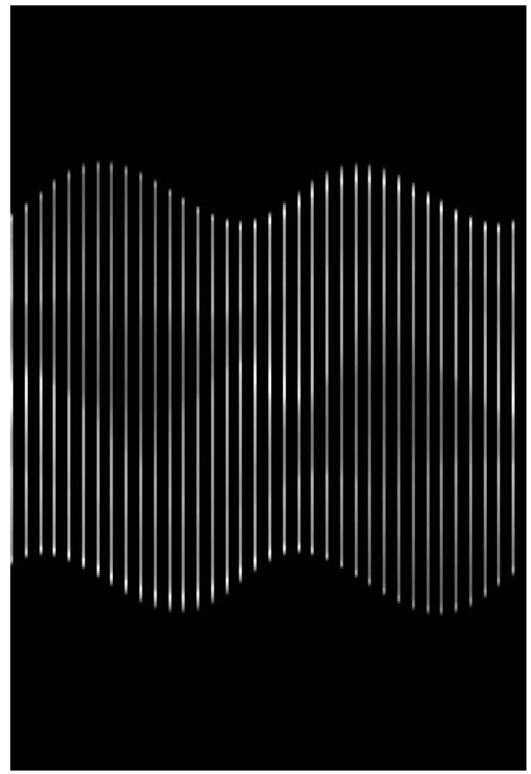
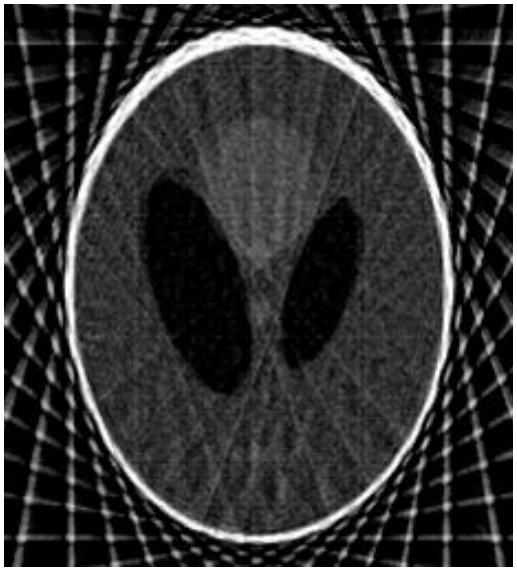


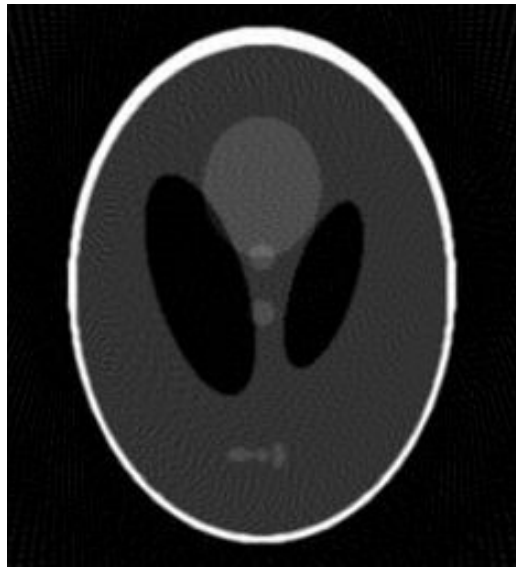
Figure : The sinogram data for 180 views and 36 views. To compare them, missing columns are replaced by 0.

2D Reconstruction Results (Without Noise)

FBP 36 views (RMSE = 50.8394)



FBP 180 views (RMSE = 14.1995)



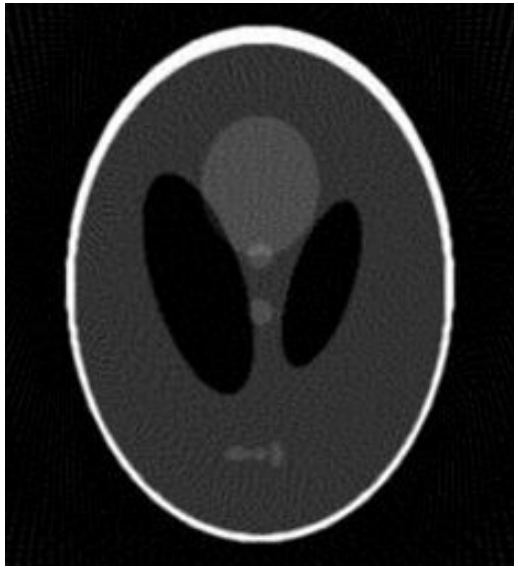
FBP 360 views (RMSE = 12.6068)

EM+TV 36 views (RMSE = 2.3789)

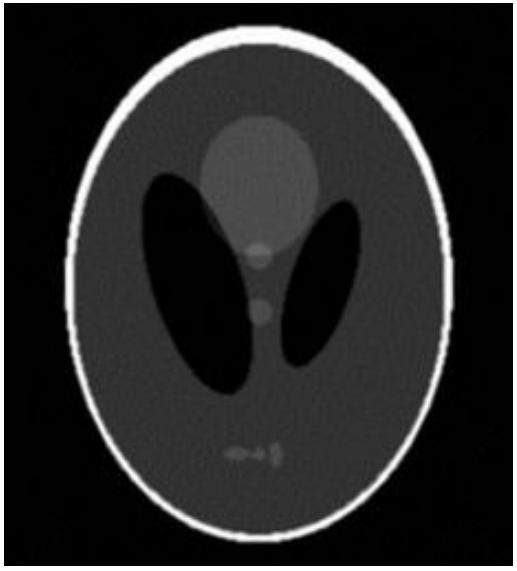
2D Reconstruction Results (With Noise)

FBP 36 views (RMSE = 51.1003)

FBP 180 views (RMSE = 14.3698)



FBP 360 views (RMSE = 12.7039)

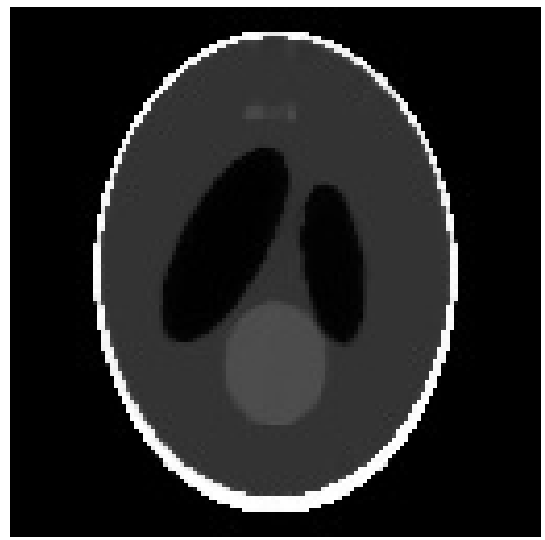


EM+TV 36 views (RMSE = 3.0868)

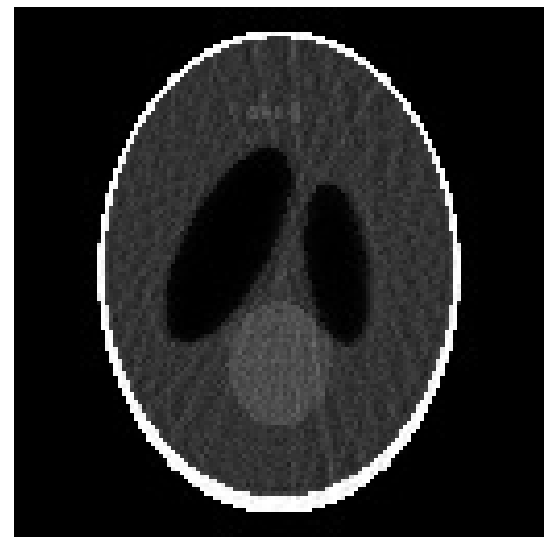
3D Reconstruction Results (3D Phantom)

original

EM+TV



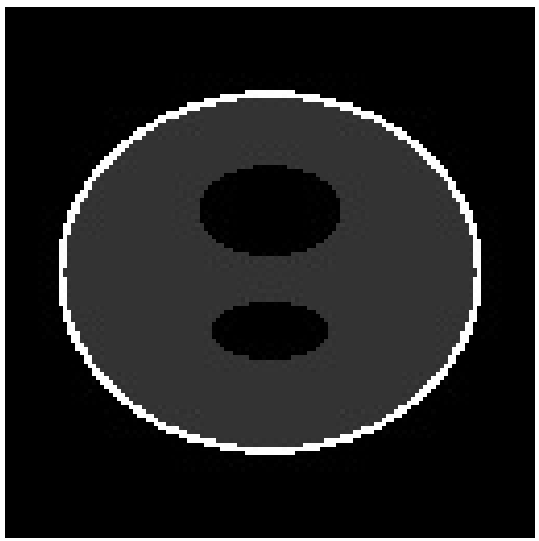
EM



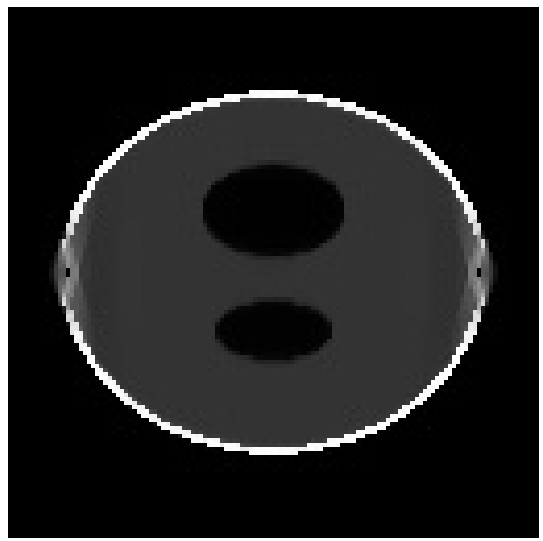
The middle slice in the z-direction.

3D Reconstruction Results (3D Phantom)

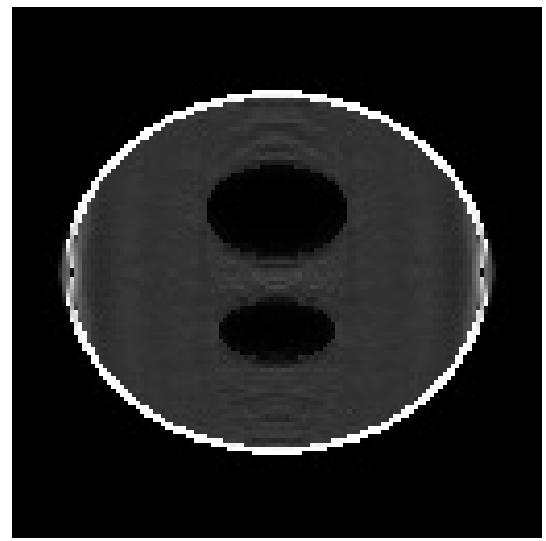
original



EM+TV



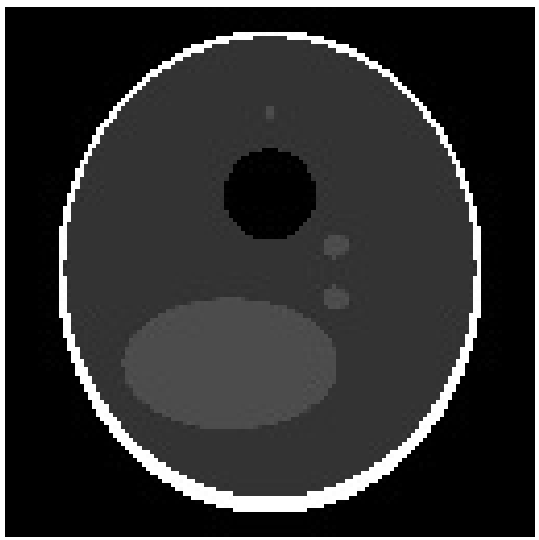
EM



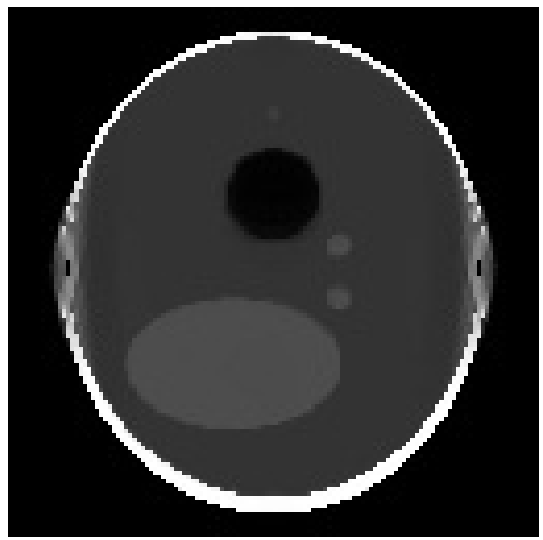
The middle slice in the y-direction.

3D Reconstruction Results (3D Phantom)

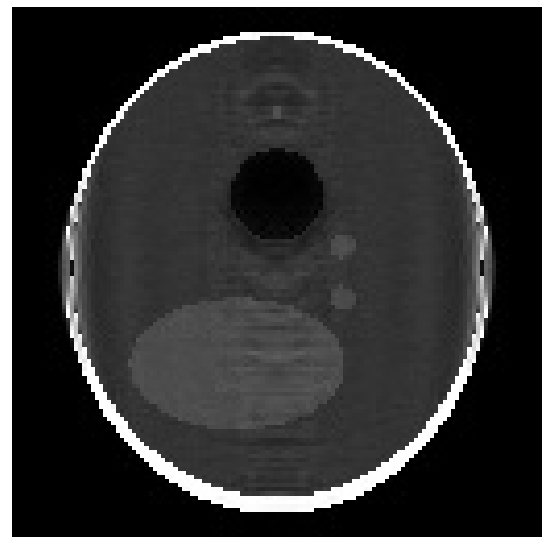
original



EM+TV



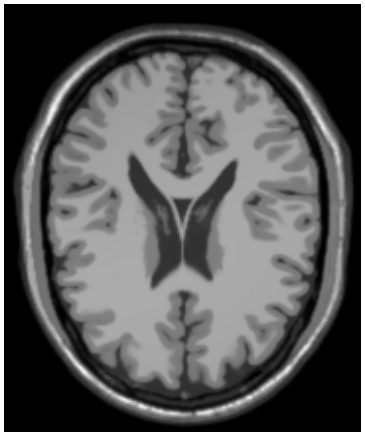
EM



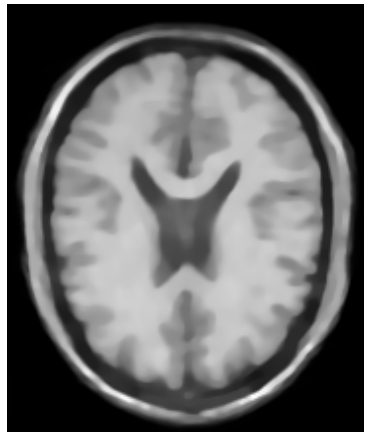
The middle slice in the x-direction.

3D Reconstruction Results using EMTV (3D MRI Brain Atlas)

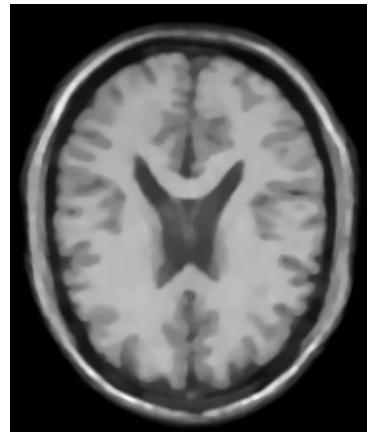
original



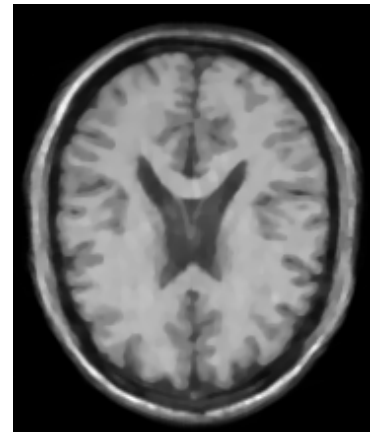
36 views(6.710)



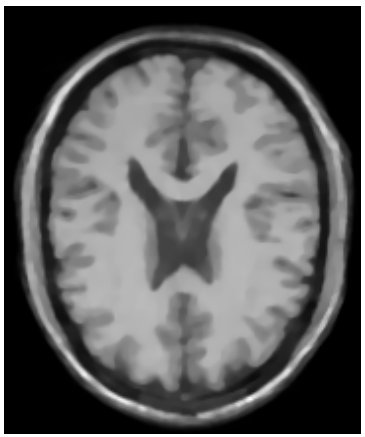
36 views(5.466)



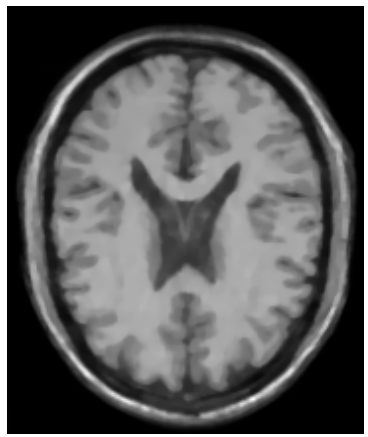
36 views(5.127)



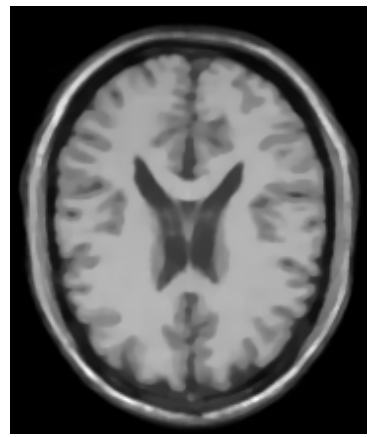
40 views(5.430)



40 views(5.080)



50 views(4.353)



50 views(3.881)