Sparsity regularized image reconstruction

QDNE 2014

Alfred Hero

University of Michigan

July 21, 2014

Outline of Presentation

Acknowledgements: sponsors

Relevant current sponsors

- ARO MURI (Value of Information): "Value-Centered Information Theory For Adaptive Learning, Inference, Tracking and Exploitation," A. Hero, PI.
- AFOSR: "Sample-starved large scale network analysis," A. Hero, PI
- NSF: "Data adaptive learning and anomaly detection," C. Scott, PI

Relevant past sponsors

- ARO: "Sparse indexing and retrieval," A. Hero, PI
- USAF/AFMC: "Data fusion for detection of anomalous structures," A. Hero, PI
- DARPA MOSAIC (Molecular Observation, Spectroscopy and Imaging using Cantilevers): "Single spin magnetic resonance force microscopy." D. Rugar, PI
- ARO MURI: "Advancement of Magnetic Resonance Force Microscopy to Single Nuclear Spin Detection." J. Sidles, PI

Acknowledgements: students and collaborators in sparsity

Current students

- Yu Hui Chen, U Michigan
- Hamed Firouzi, U Michigan
- Kristjan Greenwald, U Michigan
- Jimmy Li, U Michigan
- Kevin Moon, U Michigan

Former students

- Jeff Calder, UC Berkeley
- Kevin Carter, MIT LL
- Kyle Herrity, Freescale Austin
- Joyce Liu, UC Berkeley
- Se Un Park, Schlumberger
- Raghu Rangarajan, Cisco
- Paul Shearer, U Michigan
- Kumar Sricharan, Xerox Parc
- Michael Ting, Seagate
- Ted Tsiligkaridis, MIT LL

Current and former postdocs

- Nicolas Dobigeon, U Toulouse
- Alex Kulesza, U Michigan
- Goran Marjanovic, U Michigan
- Raviv Raich, Oregon State
- Dennis Wei, IBM Watson
- Ami Wiesel, Hebrew U

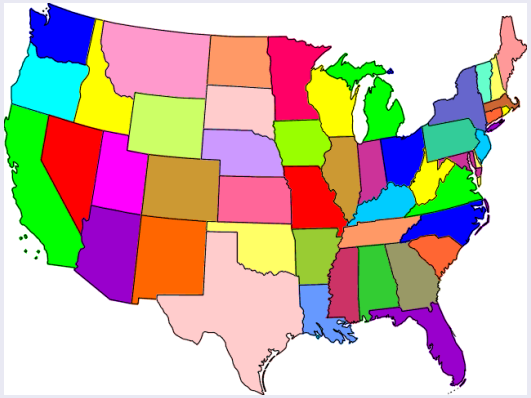
Collaborators

- Marc DeGraef, CMU
- Anna Gilbert, U Michigan
- Jeff Fessler, U Michigan
- Bala Rajaratnam, Stanford
- Indika Rajapakse, U Michigan
- Dan Rugar, IBM Almaden
- John Sidles, U. Washington
- J.-Y. Tourneret, U Toulouse
- Shuheng Zhou, U Michigan

Sparse images?

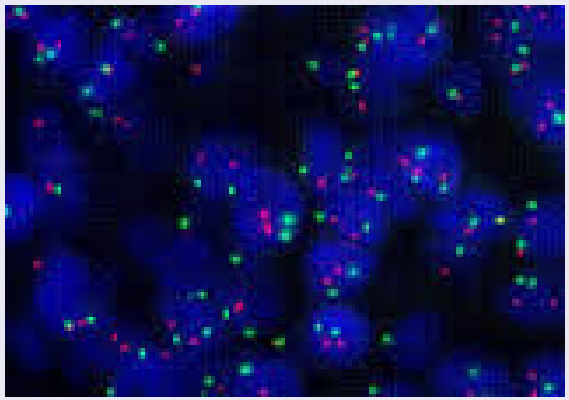
Astronomy image

Map image

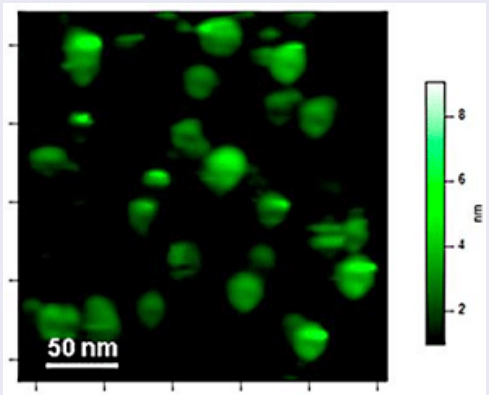


NYC Image

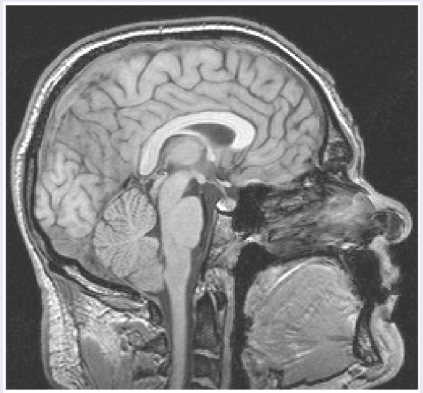
FISH HER2 Image



Protein molecules



MRI scan image



1. Sparse signals and images

Sparsity has been exploited in signal/image processing, deconvolution, and statistics for decades.

S. Mallat, A theory for multiresolution signal decomposition, *IEEE PAMI*, 1989
H.L. Taylor, S.C. Banks, and J.F. McCoy. Deconvolution with the 1 norm, *Geophysics* 1979
R. Tibshirani, Regression shrinkage and selection via the lasso. *JRSS B*, 1996

Applications of sparsity

- Representation: some data is naturally and intrinsically sparse
- Regularization: sparsity narrows down the solution space in inverse problems
- Visualization: dimensionality reduction, compression

Mathematical definition of sparsity

Define the counting measure (support function or "l_0 norm") of a vector $\mathbf{y} = [y_1, \ldots, y_n]^T \in \mathbb{R}^n$

$$\|\mathbf{y}\|_0 = \mathtt{nnz}(\mathbf{y}) = \sum_{i=1}^{n} (y_i)^0$$

Sparsity: a vector $\mathbf{y} \in \mathbb{R}^n$ is sparse if most of its elements are zero:

$$\|\mathbf{y}\|_0 = d \ll n$$

Compressibility: A discrete signal $\mathbf{x} \in \mathbb{R}^n$ is compressible with respect to a basis $\mathbf{Q} \in \mathbb{R}^{n,n}$ if $\tilde{\mathbf{x}} = \mathbf{Q}\mathbf{x}$ is sparse:

$$\|\mathbf{Q}\mathbf{x}\|_0 = d \ll n$$

A basis $\mathbf{Q}$ is an invertible matrix mapping $\mathbb{R}^n$ to $\mathbb{R}^n$.

Examples of bases

- $\mathbf{Q} = \mathbf{I}$: natural basis

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{Q}\mathbf{x} = \mathbf{x}$$

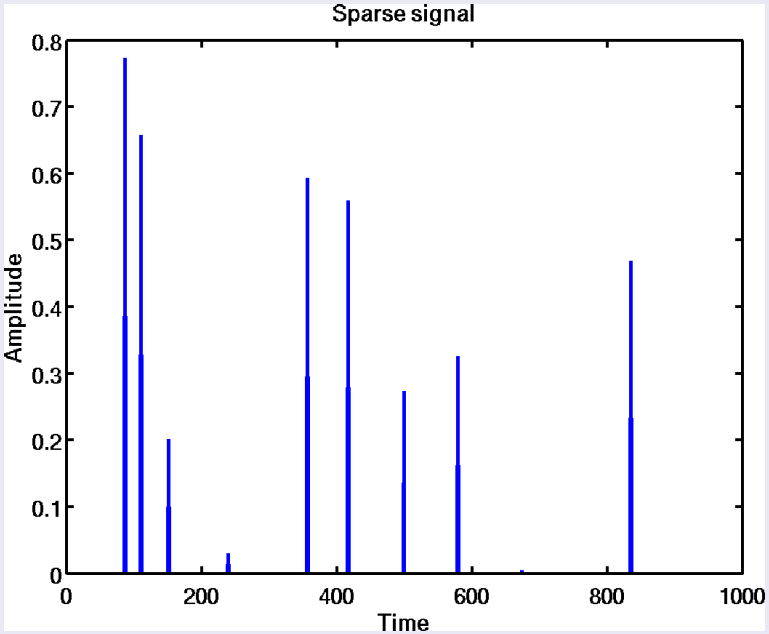
- $\mathbf{Q} = \mathbf{R}$, $\mathbf{R}\mathbf{x} = \mathbf{x} - z^{-1}\mathbf{x}$: differencing basis

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 1 & -1 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}, \qquad \mathbf{Q}\mathbf{x} = \begin{bmatrix} x_1 - x_2 \\ \vdots \\ x_{n-1} - x_n \\ x_n \end{bmatrix}$$

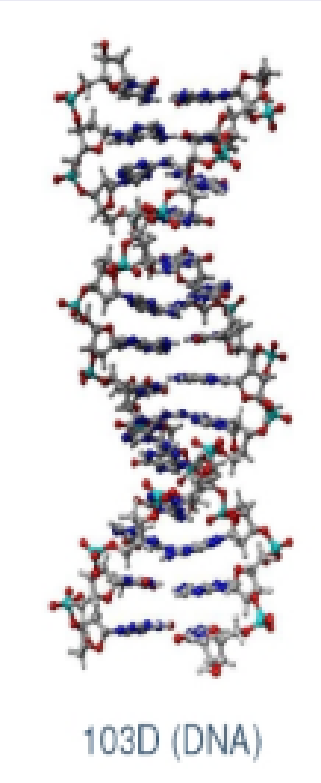
- $\mathbf{Q} = \mathbf{E}$: Fourier basis
- $\mathbf{Q} = \mathbf{W}$: Wavelet basis

Examples of sparse signals and images: $\mathbf{Q} = \mathbf{I}$

Sparse signal in $\mathbf{Q} = \mathbf{I}$

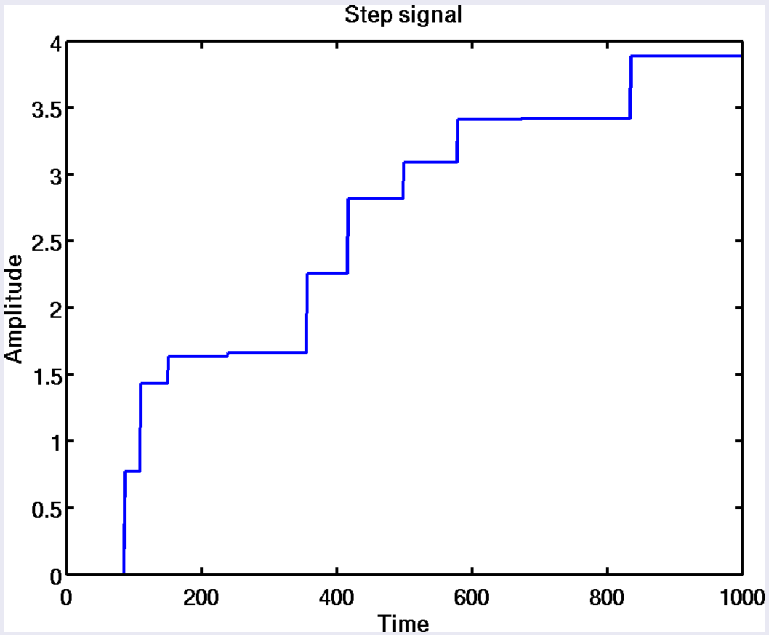


Sparse image in $\mathbf{Q} = \mathbf{I}$

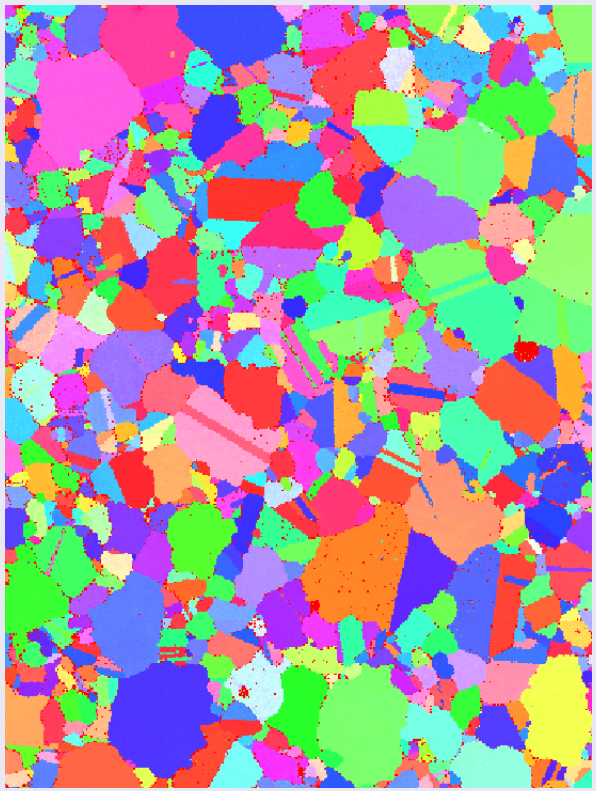


Examples of sparse signals and images: $\mathbf{Q} = \mathbf{R}$

Sparse signal in $\mathbf{Q} = \mathbf{R}$

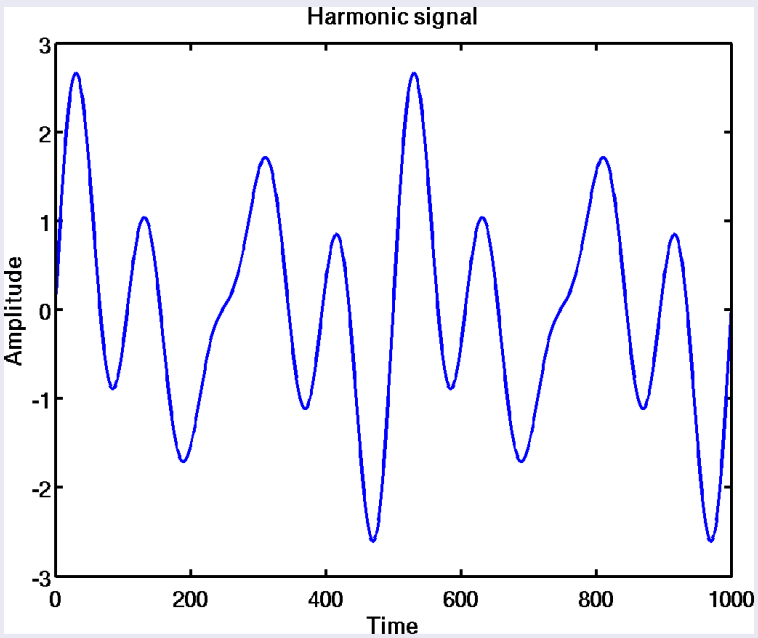


Sparse image in $\mathbf{Q} = \mathbf{R}$

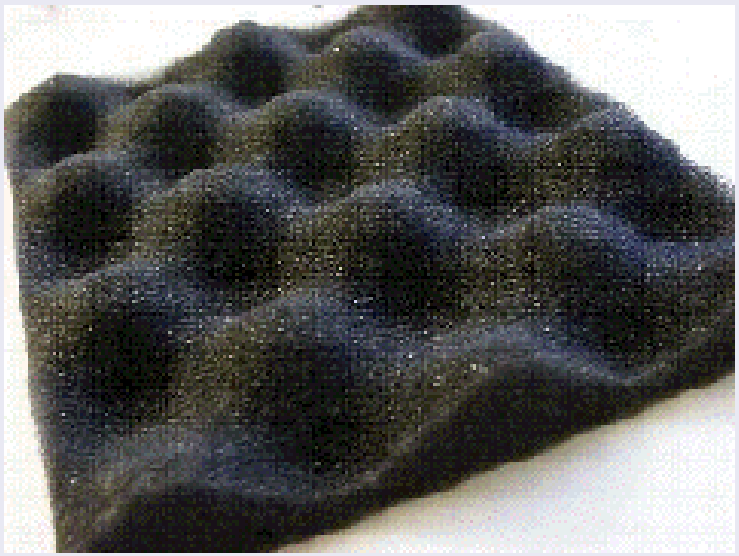


Examples of sparse signals and images: $\mathbf{Q} = \mathbf{E}$

Sparse signal in $\mathbf{Q} = \mathbf{E}$



Sparse image in $\mathbf{Q} = \mathbf{E}$



2. Image reconstruction

Blurring and *point spread* is intrinsic to image formation

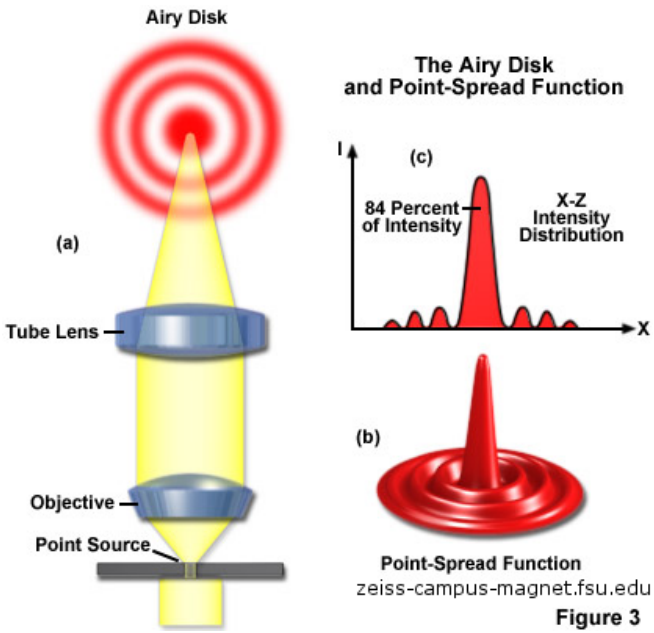


Figure 3

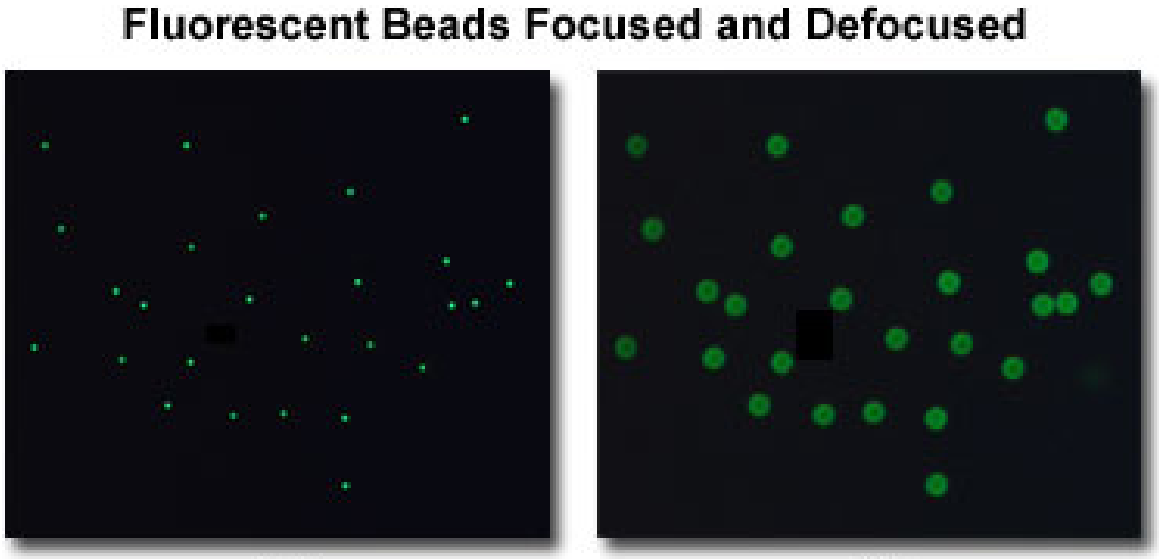


Figure 4

Deblurring is an *inverse problem*

Inverse problem: mathematical formulation

The observation equation is:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v},$$

- $\mathbf{y}$: $m \times 1$ output of instrument (blurred image, projections, ...)
- $\mathbf{x}$: $n \times 1$ lexicographically vectorized source or image
- $\mathbf{v}$: $m \times 1$ dimensional noise vector.
- $\mathbf{H}$: $m \times n$ matrix (forward operator, psf, ...)
- Sum of squared errors (SSE) prescribes goodness of *data fit*
  $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 = (\mathbf{y} - \mathbf{H}\mathbf{x})^T(\mathbf{y} - \mathbf{H}\mathbf{x})$

Solving the inverse problem

Estimation of $\mathbf{x}$ under positivity, sparsity or smoothness constraints.

Inverse problem: mathematical formulation

The observation equation is:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v},$$

- $\mathbf{y}$: $m \times 1$ output of instrument (blurred image, projections, ...)
- $\mathbf{x}$: $n \times 1$ lexicographically vectorized source or image
- $\mathbf{v}$: $m \times 1$ dimensional noise vector.
- $\mathbf{H}$: $m \times n$ matrix (forward operator, psf, ...)
- Sum of squared errors (SSE) prescribes goodness of *data fit*
  $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 = (\mathbf{y} - \mathbf{H}\mathbf{x})^T(\mathbf{y} - \mathbf{H}\mathbf{x})$

Solving the inverse problem

Estimation of $\mathbf{x}$ under positivity, sparsity or smoothness constraints.

Solving the constrained inverse problem

The observation equation is:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v},$$

The constraints are:

$$\mathbf{x} \in \mathcal{C}$$

$\mathcal{C}$ is a constraint set, e.g.,

$$\mathcal{C} = \begin{cases} \|\mathbf{Q}\mathbf{x}\|_0 \leq k, & \text{sparsity in basis } \mathbf{Q} \\ \|\mathbf{P}\mathbf{x}\|_2^2 \leq \beta, & \text{smoothness in basis } \mathbf{P} \\ \mathbf{x} \geq 0 & \text{non-negativity} \end{cases}$$

Reconstruction is solution to constrained optimization problem

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \quad \mathbf{x} \in \mathcal{C}$$

May also be written more compactly as

$$\min_{\mathbf{x} \in \mathcal{C}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$$

Solving the inverse problem: unconstrained case

First consider unconstrained least squares (LS):

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$$

Direct solution

If $m > n$ and $\mathbf{H}$ is full rank then complete the square

$$\begin{aligned} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 &= (\mathbf{y} - \mathbf{H}\mathbf{x})^T(\mathbf{y} - \mathbf{H}\mathbf{x}) \\ &= (\mathbf{x} - \widehat{\mathbf{x}})^T\mathbf{H}^T\mathbf{H}(\mathbf{x} - \widehat{\mathbf{x}}) + \mathbf{y}^T(\mathbf{I} - \mathbf{H}[\mathbf{H}^T\mathbf{H}]^{-1}\mathbf{H}^T)\mathbf{y} \end{aligned}$$

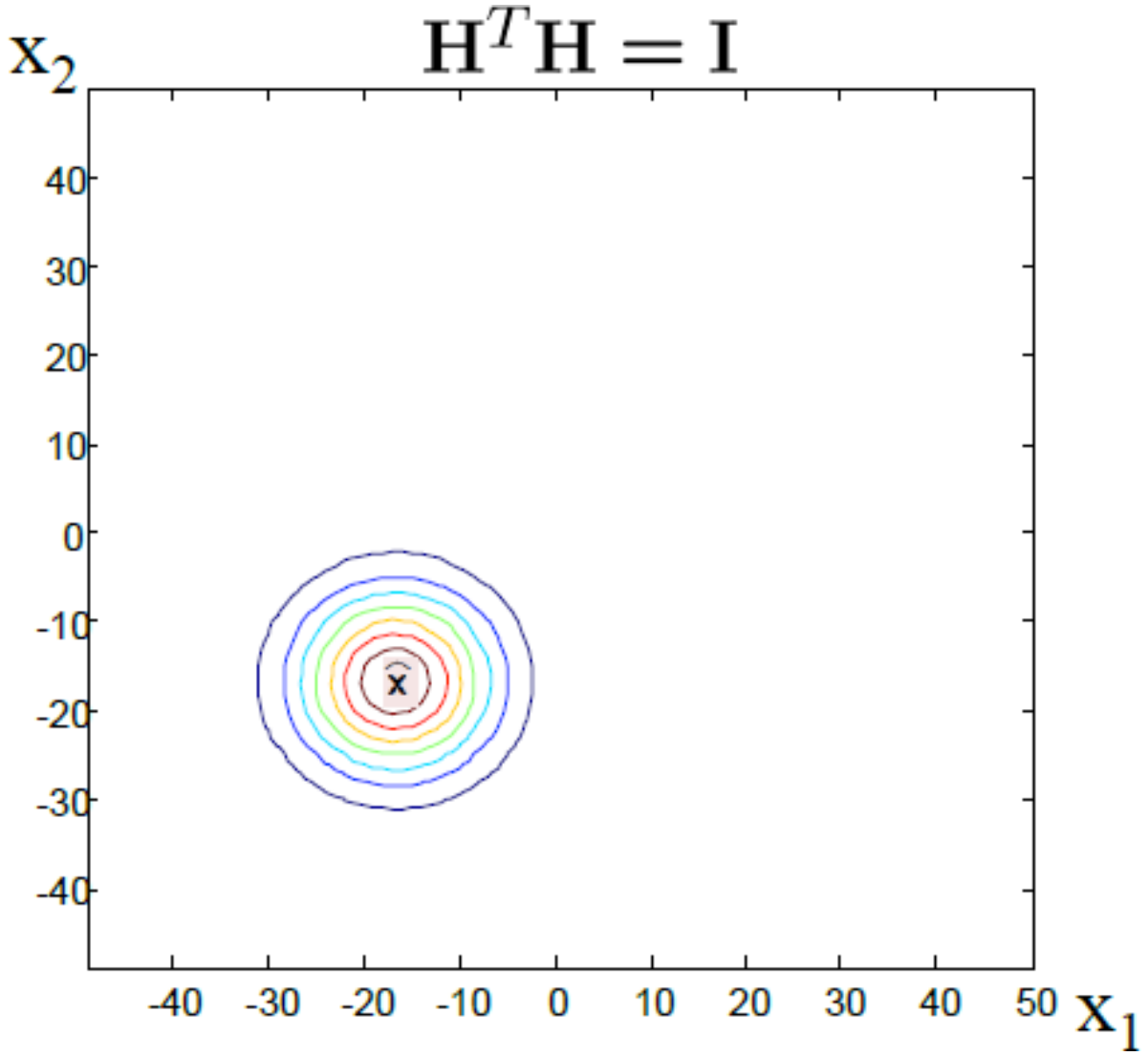
where

$$\widehat{\mathbf{x}} = [\mathbf{H}^T\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{y}$$

Hence, unconstrained LS soln is:

$$\mathrm{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 = \widehat{\mathbf{x}}$$

Unconstrained LS as minimum over SSE surface ($n = 2$)



Solving the inverse problem: Landweber iterations

Direct soln of unconstrained LS

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$$

requires inversion of $n \times n$ matrix $\mathbf{H}^T\mathbf{H}$

Landweber iterations

Apply gradient descent to unconstrained LS problem

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} + \frac{\lambda}{2}\nabla_{\mathbf{x}^{(t)}} \|\mathbf{y} - \mathbf{H}\mathbf{x}^{(t)}\|_2^2 \\ &= \mathbf{x}^{(t)} - \lambda \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(t)}) \end{aligned}$$

where $\lambda > 0$ is a user selectable parameter

Properties of Landweber iterations

- Iteration does not involve any matrix inversions
- Convergent as long as $\mathrm{mineig}(\mathbf{H}^T\mathbf{H}) > 0$ and $\lambda < 1/\mathrm{maxeig}(\mathbf{H}^T\mathbf{H})$
- Speed of convergence depends on condition number $\kappa(\mathbf{H}^T\mathbf{H})$

Solving the inverse problem: sparsity constrained

Next consider sparsity constrained problem (no loss in assuming $\mathbf{Q} = \mathbf{I}$)

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \quad \|\mathbf{x}\|_0 \leq k$$

Direct (variable selection) solution

Given n and $k \leq m$, consider $\|\mathbf{x}\|_0 = k$ with $k = \rho n$.
There are $\binom{n}{k}$ possible patterns of $n - k$ zero entries of $\mathbf{x}$.

$$\mathbf{x} = [0, \ldots, 0, x_{n-k+1}, \ldots, x_n], \quad \ldots \quad \mathbf{x} = [x_1, \ldots, x_k, 0, \ldots, 0]$$

Fix one of these patterns (the i-th zero pattern).
Then: $\mathbf{H}\mathbf{x} = \mathbf{H}_i\mathbf{z}$, where $\mathbf{z} \in \mathbb{R}^k$ and $\mathbf{H}_i \in \mathbb{R}^{m \times k}$.
The optimal reconstruction for this pattern is

$$\hat{\mathbf{z}}_i = \text{argmin}_{\mathbf{z}} \|\mathbf{y} - \mathbf{H}_i\mathbf{z}\|_2^2 = [\mathbf{H}_i^T\mathbf{H}_i]^{-1}\mathbf{H}_i^T\mathbf{y}$$

Solving the inverse problem

Sparsity constrained problem

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \quad \|\mathbf{x}\|_0 \leq k$$

Exhaustive variable selection strategy

1. Generate all possible solutions $\hat{\mathbf{z}}_i$, $i = 1, \ldots, \binom{n}{k}$
2. Expand $\hat{\mathbf{z}}_i$ into non-zero entries of $\hat{\mathbf{x}}_i$
3. Evaluate $\text{SSE}_i = \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_i\|_2^2$ for each solution
4. The global soln $\hat{\mathbf{x}}$ is the $\hat{\mathbf{x}}_i$ with minimum SSE

Issue: This variable selection solution strategy is impractical

Number of patterns that need to be considered for any k

$$\binom{n}{k} = \binom{n}{\rho n} \approx \frac{e^{-nh(\rho)}}{\sqrt{2\pi\rho(1-\rho)n}}$$

Exponential complexity :(

Solving the inverse problem

Sparsity constrained problem

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \quad \|\mathbf{x}\|_0 \leq k$$

Exhaustive variable selection strategy

1. Generate all possible solutions $\hat{\mathbf{z}}_i$, $i = 1, \ldots, \binom{n}{k}$
2. Expand $\hat{\mathbf{z}}_i$ into non-zero entries of $\hat{\mathbf{x}}_i$
3. Evaluate $\text{SSE}_i = \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_i\|_2^2$ for each solution
4. The global soln $\hat{\mathbf{x}}$ is the $\hat{\mathbf{x}}_i$ with minimum SSE

Issue: This variable selection solution strategy is impractical
Number of patterns that need to be considered for any k

$$\binom{n}{k} = \binom{n}{\rho n} \approx \frac{e^{-nh(\rho)}}{\sqrt{2\pi\rho(1-\rho)n}}$$

Exponential complexity :(

Sparsity regularized reconstruction

Sparse reconstruction algorithms fall into three categories

- Post-reconstruction thresholding
  - heuristic approach prevalent before 1990
  - simple and fast algorithms
  - no performance guarantees
- Non-smooth optimization methods
  - became prevalent in this millenium
  - iterative algorithms with tuning parameters
  - often have convergence guarantees
- Bayesian methods of inverse probability
  - became prevalent in 1980's
  - Monte Carlo algorithms based on statistical models
  - quantify uncertainty in addition to reconstructing images

3. Variational sparsity regularized reconstruction approaches

Original sparsity constrained problem

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \quad \|\mathbf{x}\|_0 \leq k \tag{1}$$

$\Rightarrow$ Consider ℓ_1 relaxation of ℓ_0 constraint

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \quad \|\mathbf{x}\|_1 \leq \delta \tag{2}$$

where $\delta > 0$ and $\|\mathbf{x}\|_1 = \sum_{j=1}^n |x_j|$ is the ℓ_1 norm.

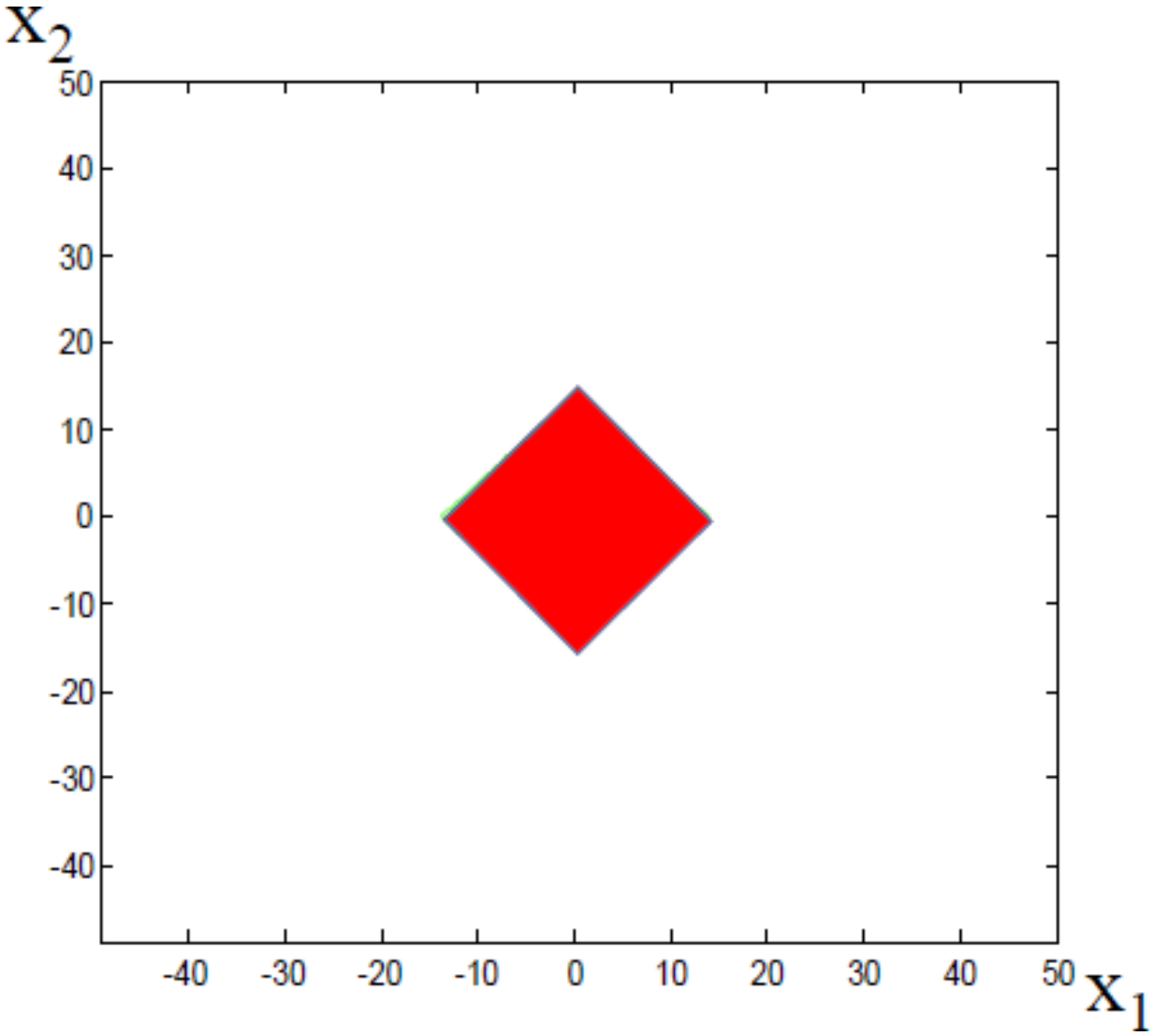
Advantages

- Minimization (2) is a convex optimization that can be solved by linear programming
- The solutions to (1) and (2) are identical if

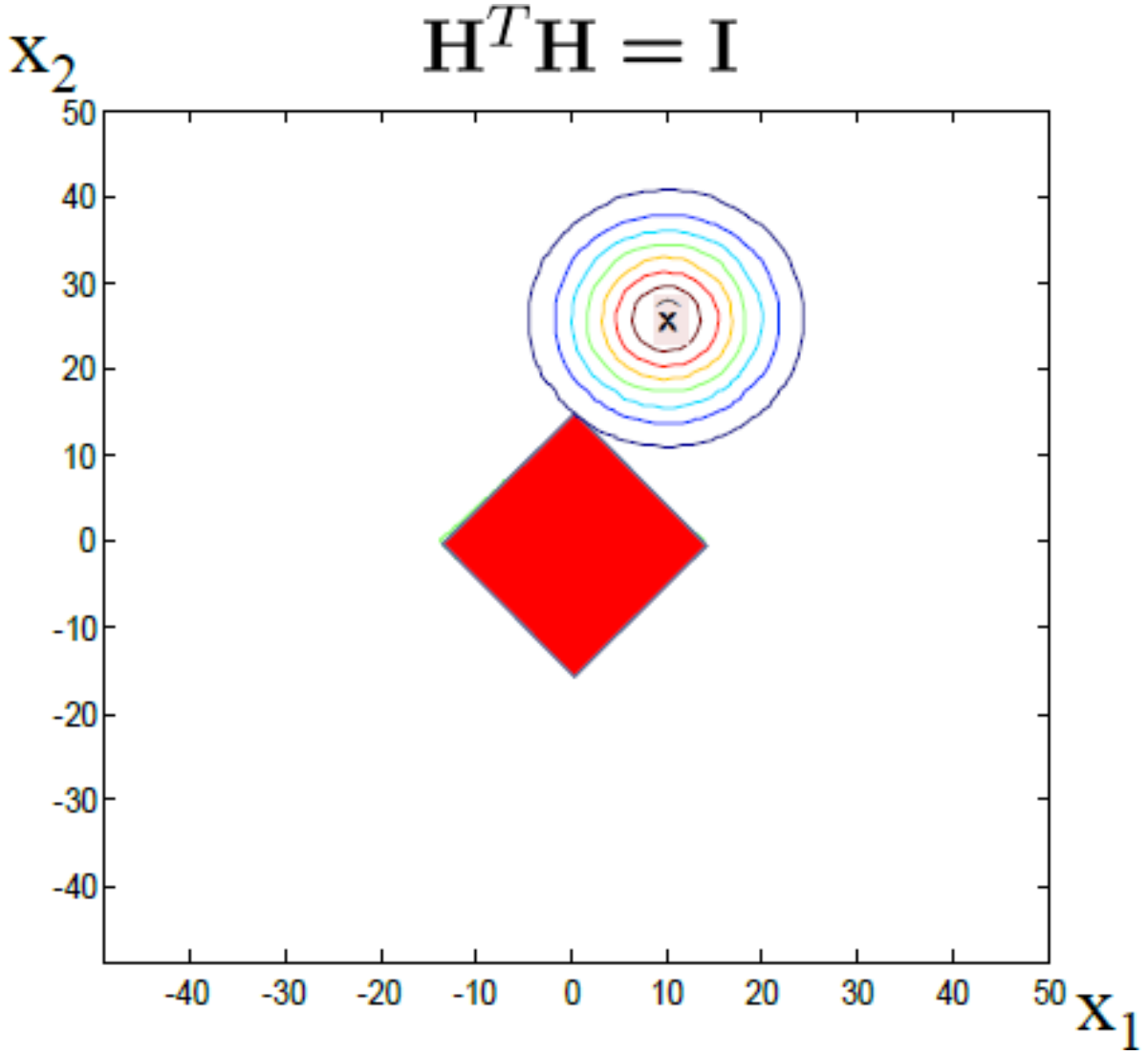
$$\mathbf{H}^T\mathbf{H} = \mathbf{I} \text{ and } n \to \infty$$

- The solutions are close under a uniform uncertainty principle (UUP): $\mathbf{H}^T\mathbf{H} \approx \mathbf{I}$ and n is large [Candes and Tao, 2006], [Donoho, 2006]

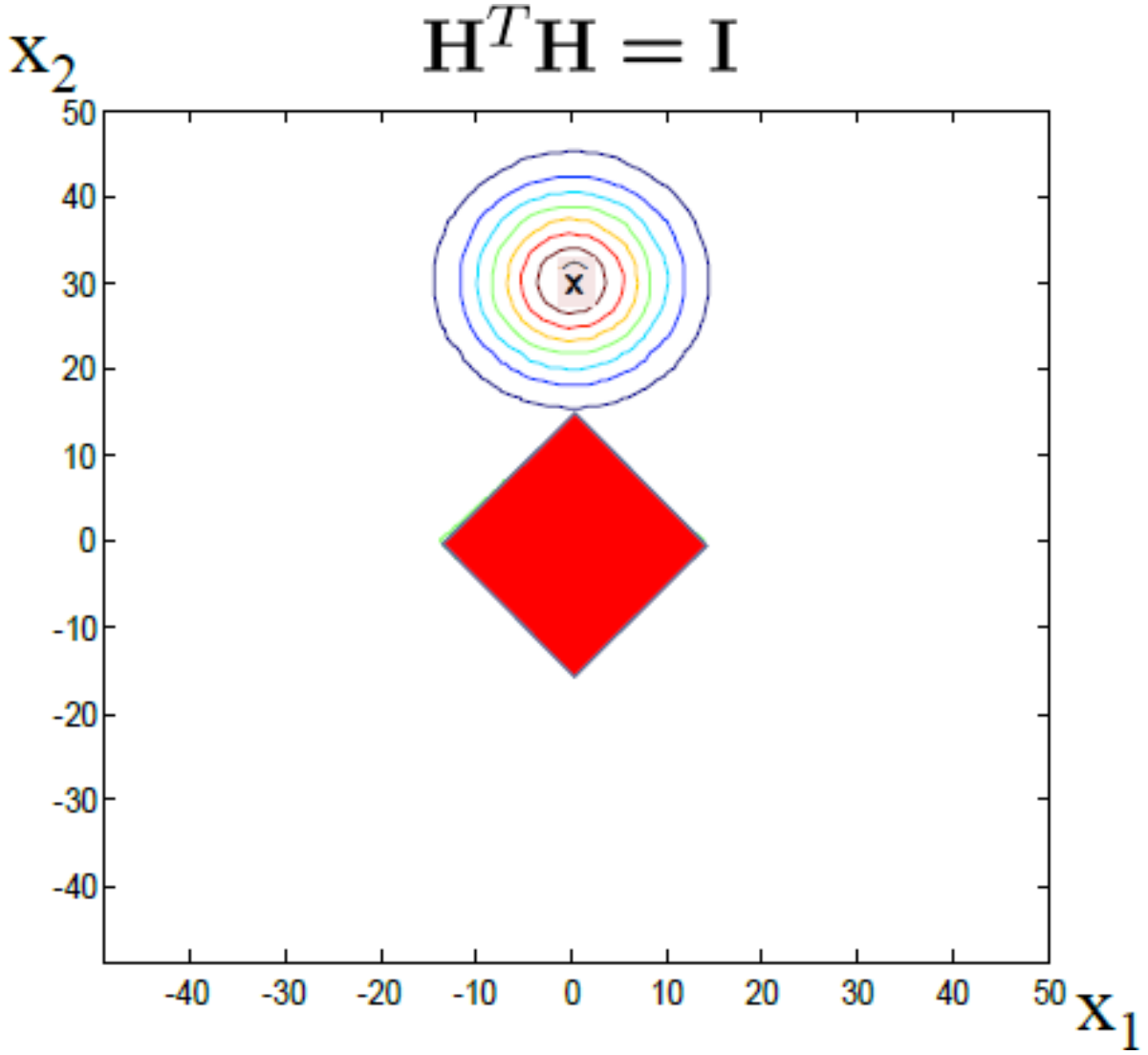
The ℓ_1 constraint region



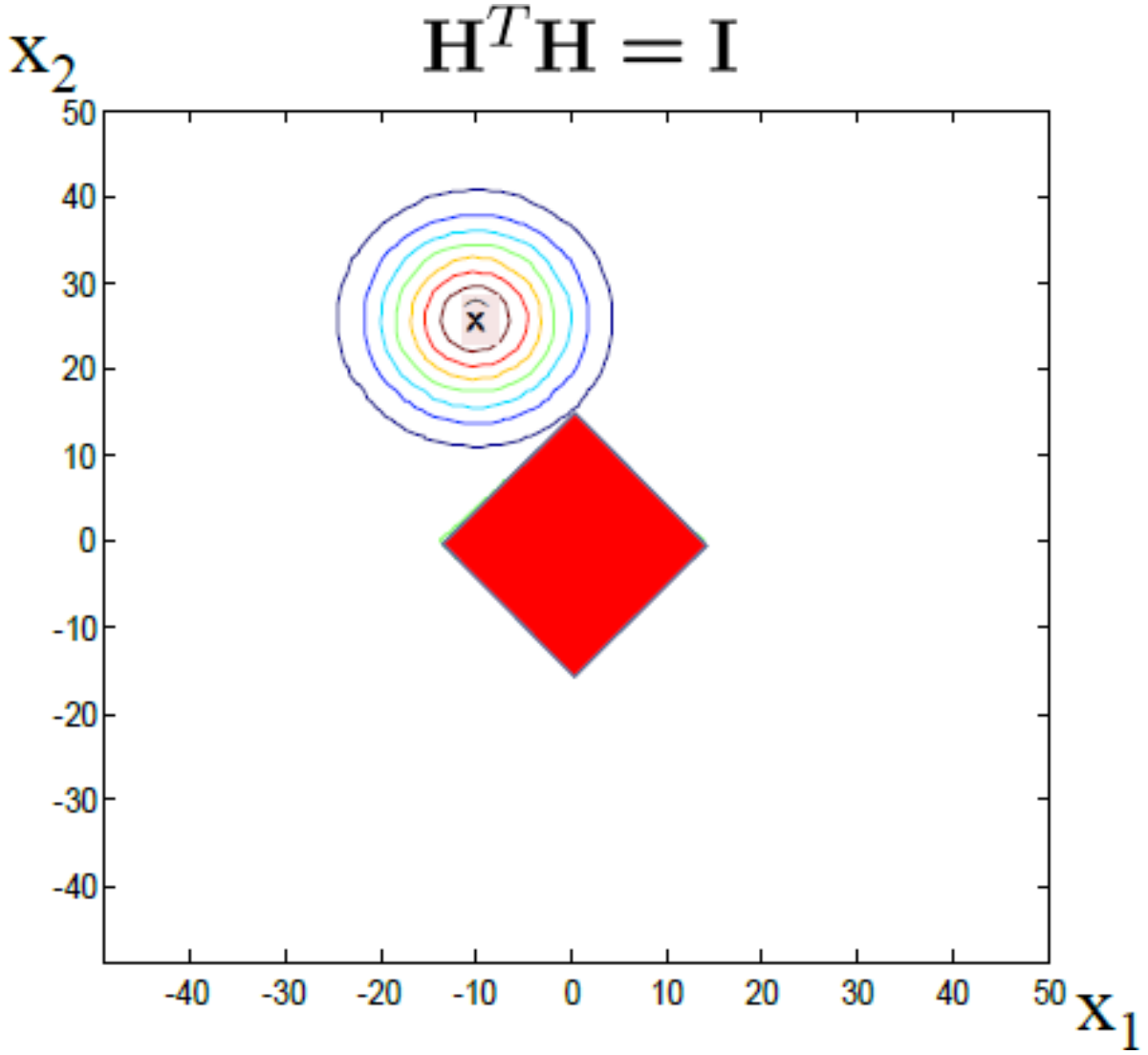
PLS sparsifies in ℓ_1 constraint region



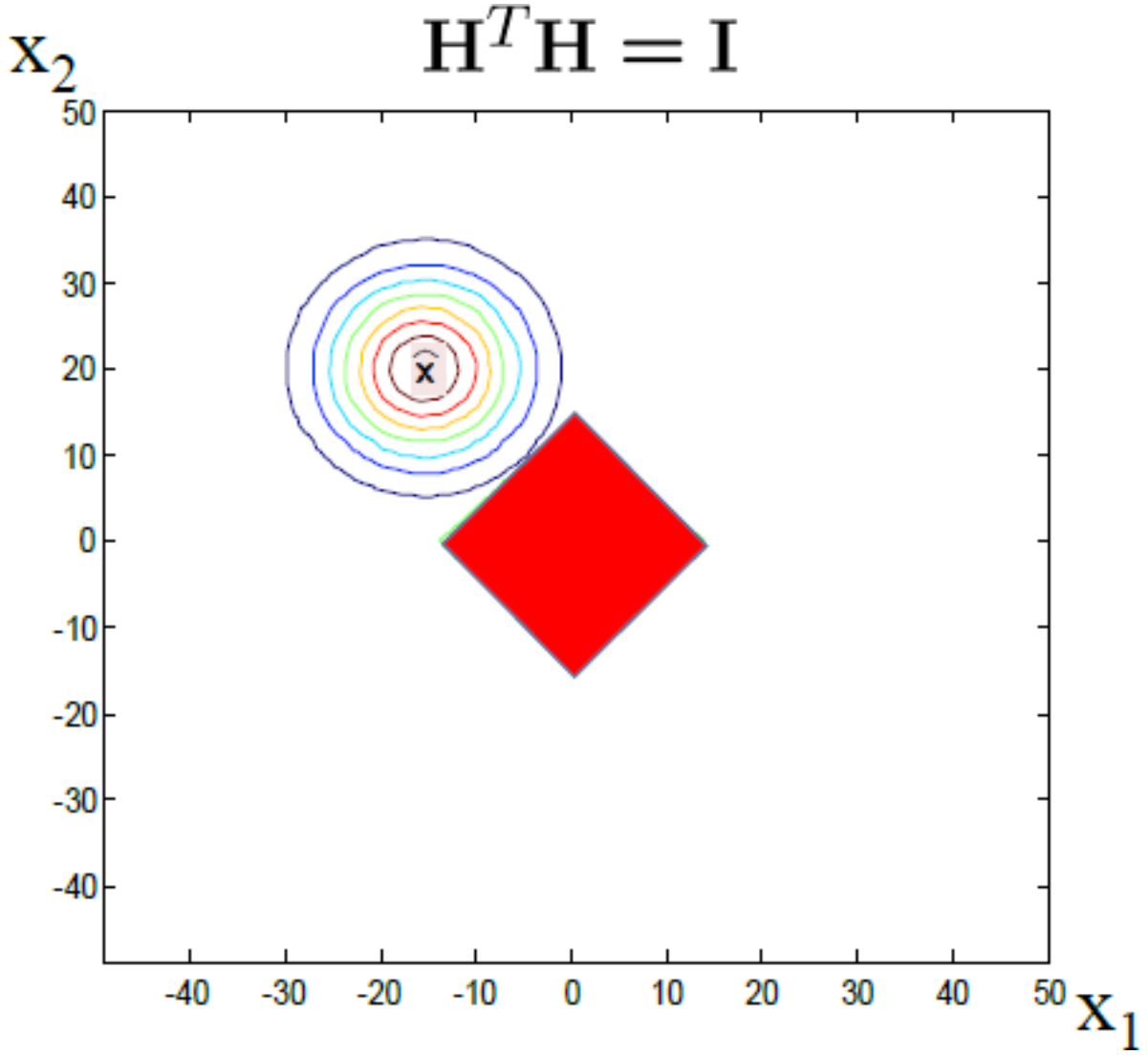
PLS sparsifies in ℓ_1 constraint region



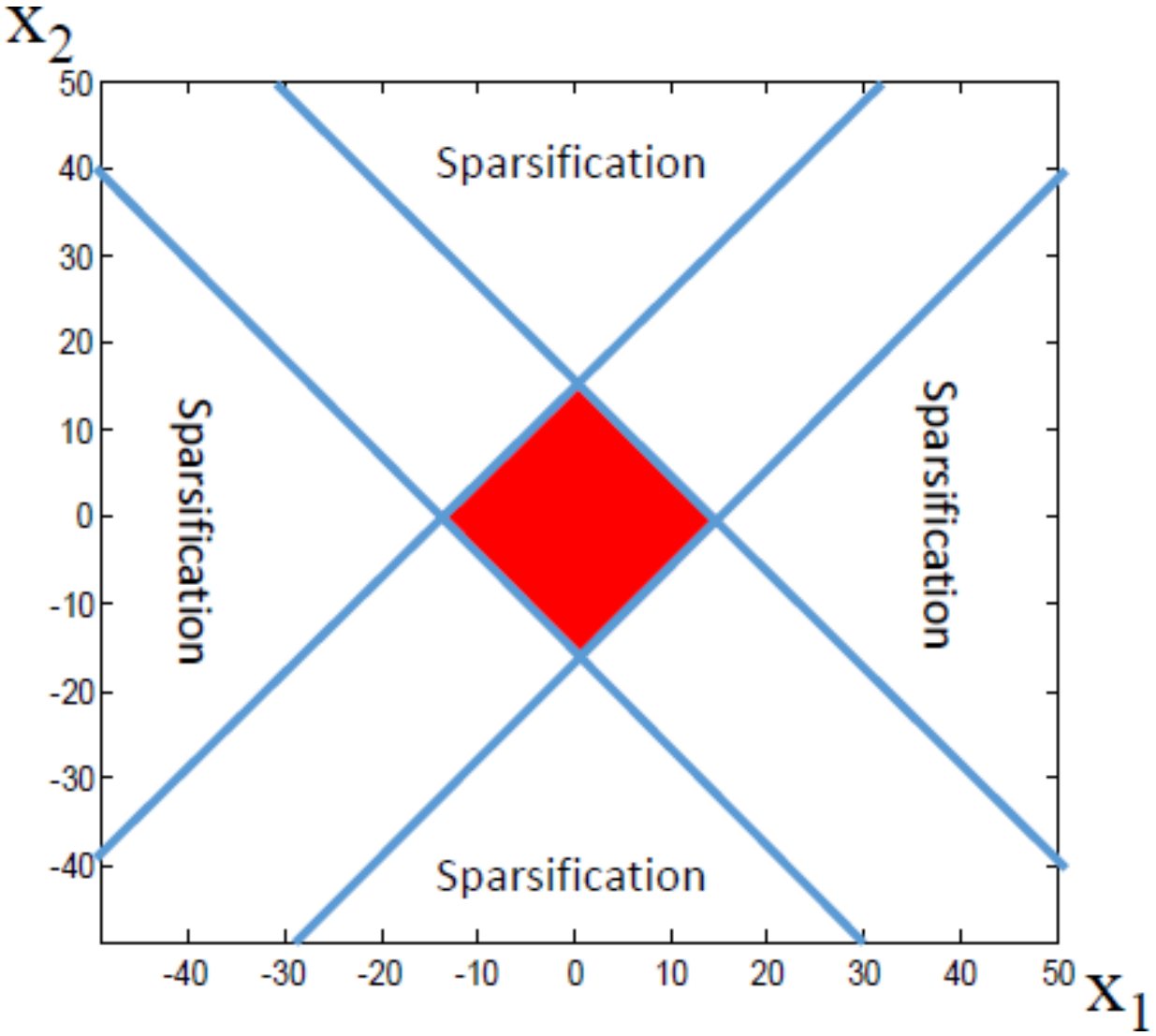
PLS sparsifies in ℓ_1 constraint region



PLS does not sparsify in ℓ_1 constraint region



The ℓ_1 PLS sparsification region



Non-smooth optimization: IST

Relaxed sparsity constrained problem

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \quad \|\mathbf{x}\|_1 \leq \delta \tag{2}$$

Iterative Soft Thresholding (IST) [Bruck 1977]

$$\mathbf{x}^{(t+1)} = \text{shrink}_{\lambda/2}\left(\mathbf{x}^{(t)} - \lambda \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(t)})\right)$$

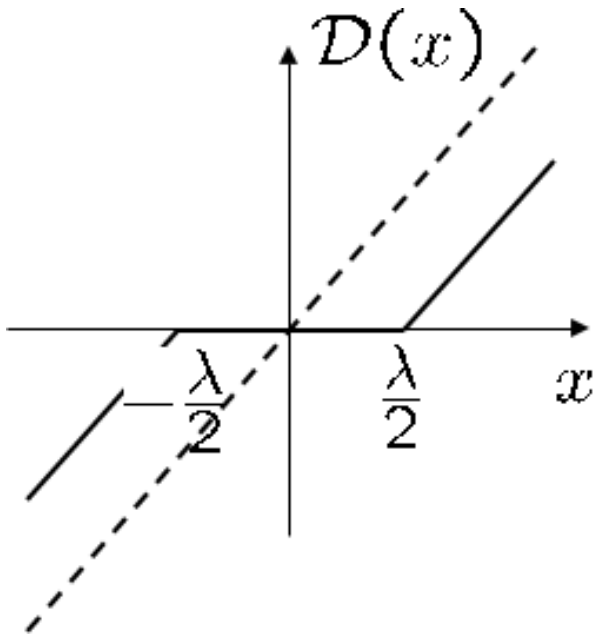
where shrink_τ is shrinkage operator

$$\text{shrink}_\tau(\mathbf{u}) = \text{sign}(\mathbf{u}) \odot \max\{|\mathbf{u}| - \tau, 0\}$$

and λ is selected to ensure that $\|\mathbf{x}\|_1 \approx \delta$. Properties

- IST is equivalent to *projected gradient algorithm* with *proximity operator* $\text{shrink}_\tau(\mathbf{u})$
- Here the proximity operator projects onto ℓ_1-ball
- Generalizes to different constraint sets and data fitting terms

Shrinkage operator



$$\mathcal{D}(x) = \text{shrink}_{\lambda/2}(x)$$

- Called soft thresholding operator in engineering and proximal operator in mathematical optimization
- Truncates small values to zero and shrinks large values
- Used in wavelet compression (JPEG) and statistics (lasso, Stein)

General projected gradient algorithm

$$\min_{\mathbf{x}} F(\mathbf{x}), \quad c(\mathbf{x}) \leq \delta \tag{3}$$

where $c(\mathbf{x})$ is constraint function and $F(\mathbf{x})$ is smooth. Examples:

- Previous quadratic ℓ_1-constrained case

$$F(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \quad c(\mathbf{x}) = \|\mathbf{x}\|_1$$

- Non-linear transport phenomena (backscatter distribution)
- Photon-limited measurements (Poisson likelihood)
- Non-gaussian heavy-tailed noise (Laplacian or Student-t likelihood)

Projected gradient algorithm [Combettes, Pesquet 10]

$$\mathbf{x}^{(t+1)} = \mathcal{P}_{\lambda/2,c}\left(\mathbf{x}^{(t)} + \frac{\lambda}{2}\nabla_x F(\mathbf{x})\right)$$

where $\mathcal{P}_{\tau,c}$ is proximity operator [Moreau 62]

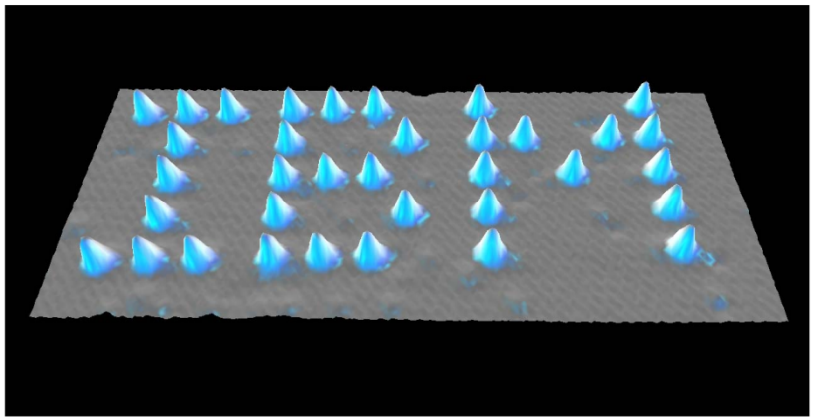
$$\mathcal{P}_{\tau,c}(\mathbf{u}) = \text{argmin}_{\mathbf{x}} \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 + \tau c(\mathbf{x})$$

Illustration of IST: nuclear spin imaging

Atomic Force and Nuclear Magnetic Resonance Modalities

- AFM: reads out deflection of cantilever scanning sample surface.
- NMR: reads out bulk spin density in a sample volume.
- NMR is volumetric, chemically selective, non-destructive but requires $10^{14} - 10^{18}$ spins/voxel.

AFM



Xenon atoms on Ni(110) - Eigler and Schweizer (1990)

NMR-MRI

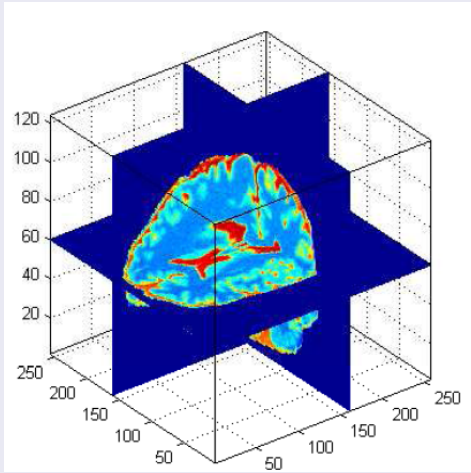
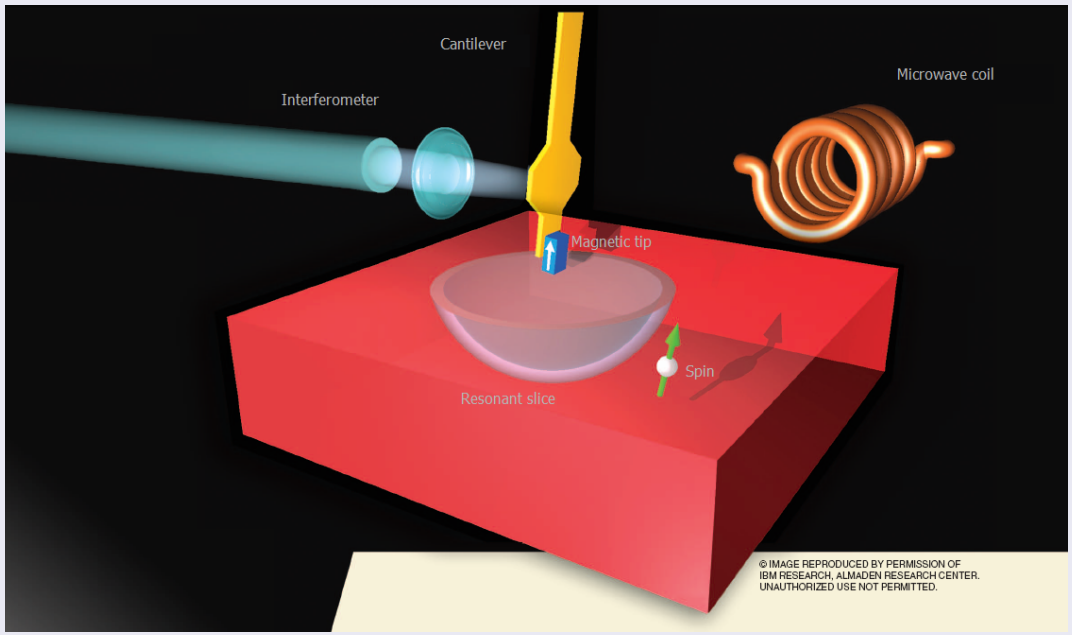


Illustration: nuclear spin imaging

Magnetic Resonance Force Microscopy (MRFM)

- chemically selective 3D volumetric single spin resolution.
- non-destructive imaging method introduced by Sidles (APL 1992),
- refined by Rugar (Nature 1992, Nature 2004, PNAS 2009).

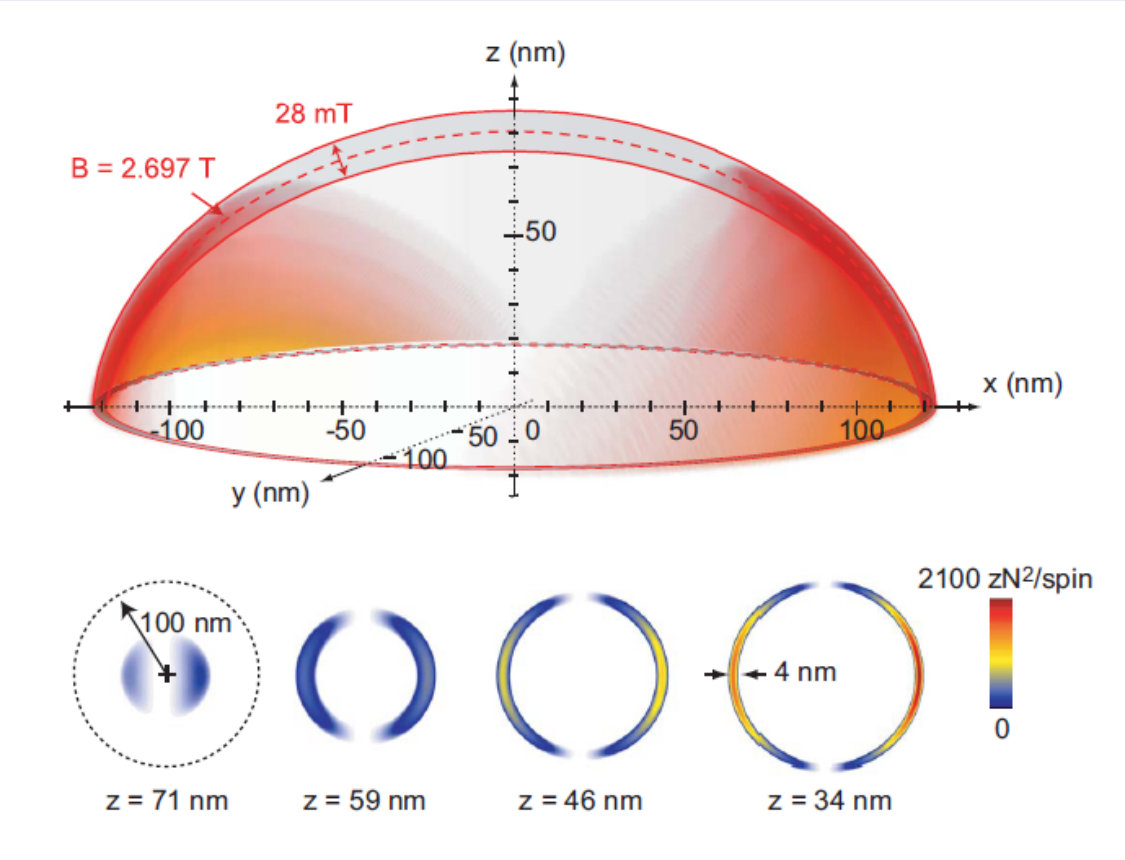
Dan Rugar's IBM single-spin MRFM instrument



The MRFM image reconstruction problem

IBM single-spin MRFM instrument

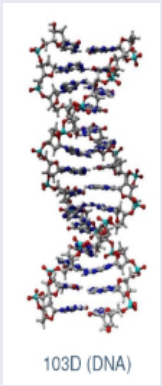
Point spread function (PSF) of the magnetic tip

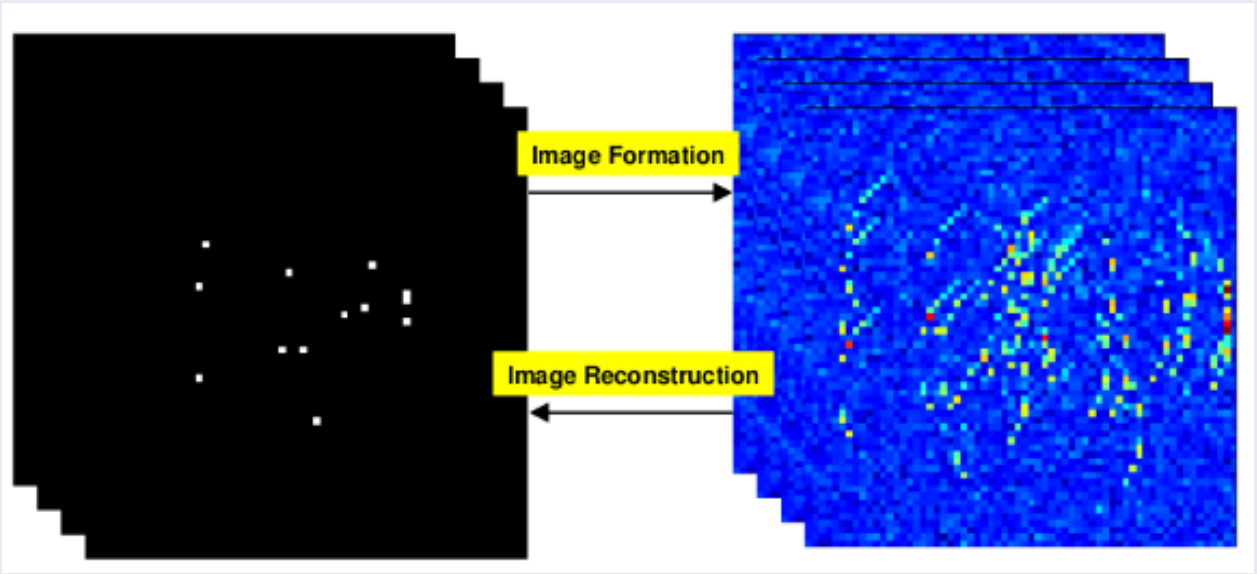


Reference: C. L. Degen *et al.*, Proc. Nat. Academy of Science, 2009.

The MRFM image reconstruction problem

3D volumetric imaging = inverse problem





- Smoothness constraints: Zügar and Rugar, J. App. Physics 1994
- Complexity constraints: Hammel,. . . Rourkes, Proc. IEEE 2003
- Sparsity constraints: Ting, Raich, H, IEEE TIP 2009
- Sparsity priors: Dobigeon, Tourneret, H, IEEE TIP 2009
- Blind sparse reconstruction: Park, Dobigeon, H, IEEE TIP 2012 and Park, Dobigeon, H IEEE TSP 2013.

Quantitative comparisons: 2D slice [Ting, Raich, H 09]

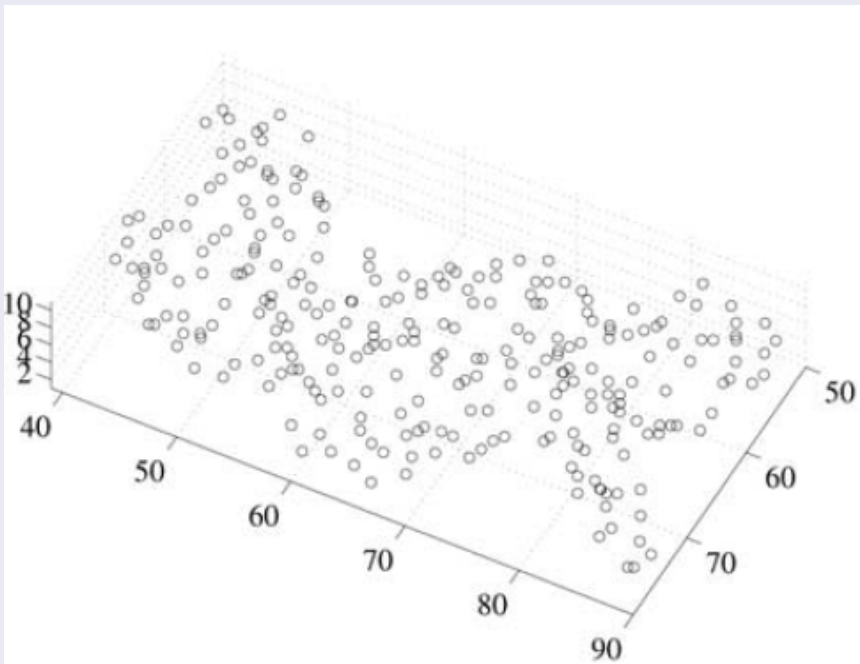
Relative performance of IST (Lasso-Sure) vs Landweber (LS)

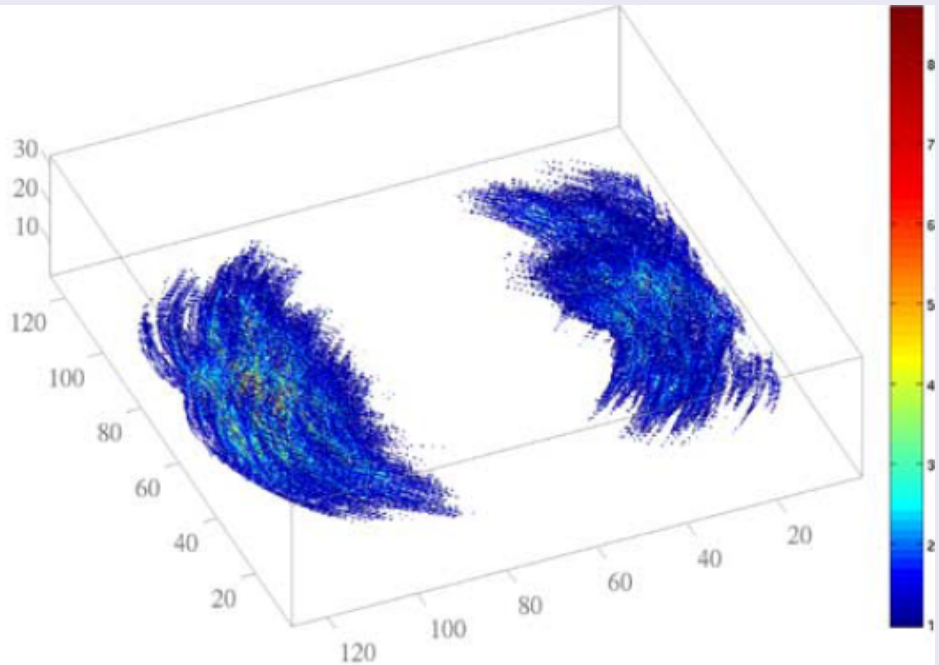
	Error criterion				
Method	$\|\underline{e}\|_0$	$\|\underline{e}\|_1$	$\|\underline{e}\|_2$	$E_d(\underline{\theta}, \hat{\underline{\theta}})$	$\|\hat{\underline{\theta}}\|_0$
SNR = 1.76 dB					
Oracular LS	27	5.71	1.55	0.56	27
LS	1024	807	31.6	977	1024
SBL	1024	28.1	3.99	72.6	1024
StOMP (CFAR)	264	4.37×10^3	558	244	257
StOMP (CFDR)	409	1.62×10^4	1.65×10^3	386	405
MAP1	27	21.2	5.21	27	0
MAP2	30.9	17.5	3.98	25.1	9.77
lasso-SURE	92.6	20.3	3.15	69.3	81.9
H-SURE	67.2	19.1	3.14	51.1	54.7

3D comparison [Ting, Raich, H 2009

$B_{ext} = 2.88 \cdot 10^4 G, B_{res} = 2 \cdot 10^4 G, G_{max} = 407 G/nm, R_0 = d = 3nm, x_{pk} = 0.05nm$

3D image formation for DNA molecule



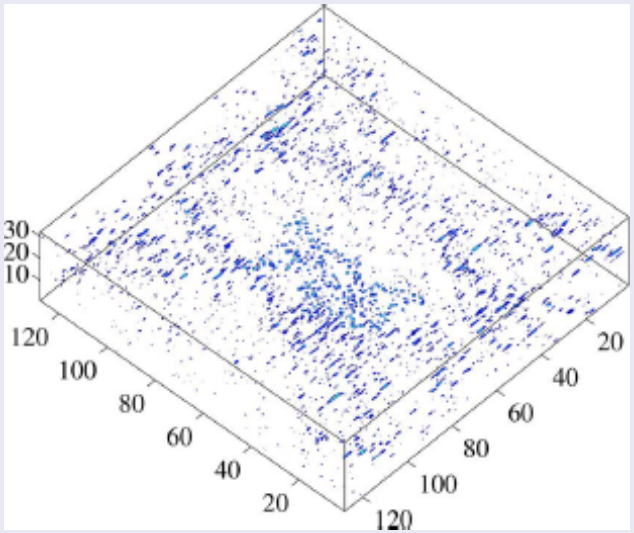


Hydrogen atom locations of 103D

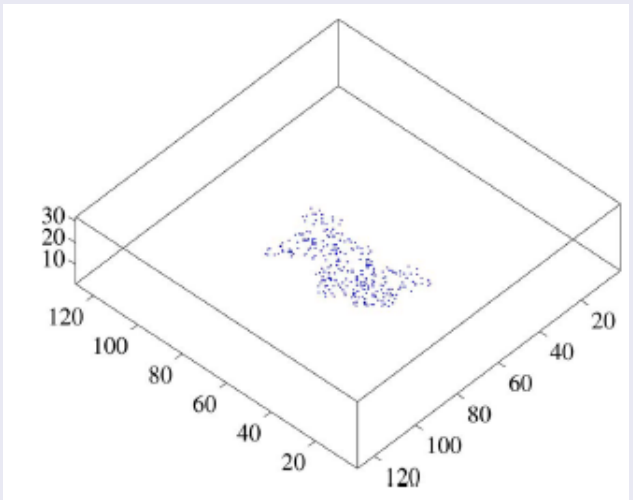
MRFM projections

IST vs Landweber reconstruction [Ting, Raich, H 2009

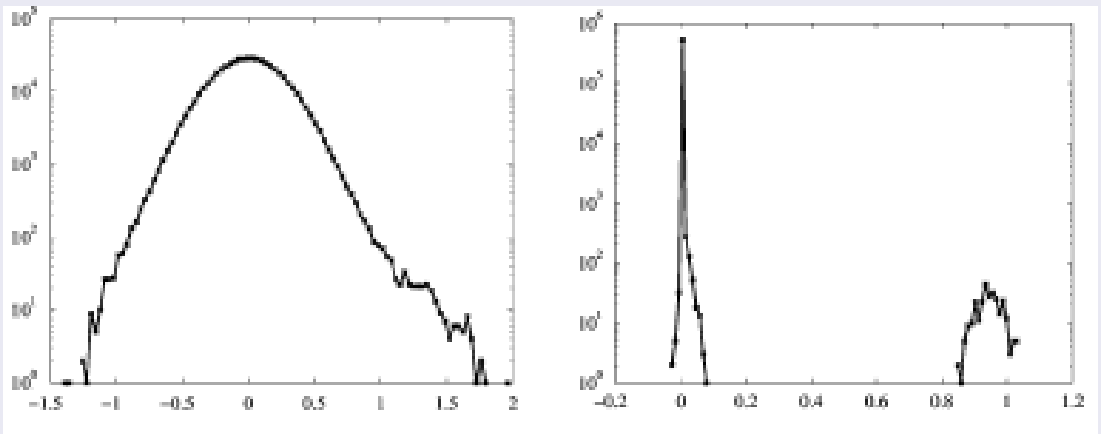
Landweber Reconstruction



IST Reconstruction

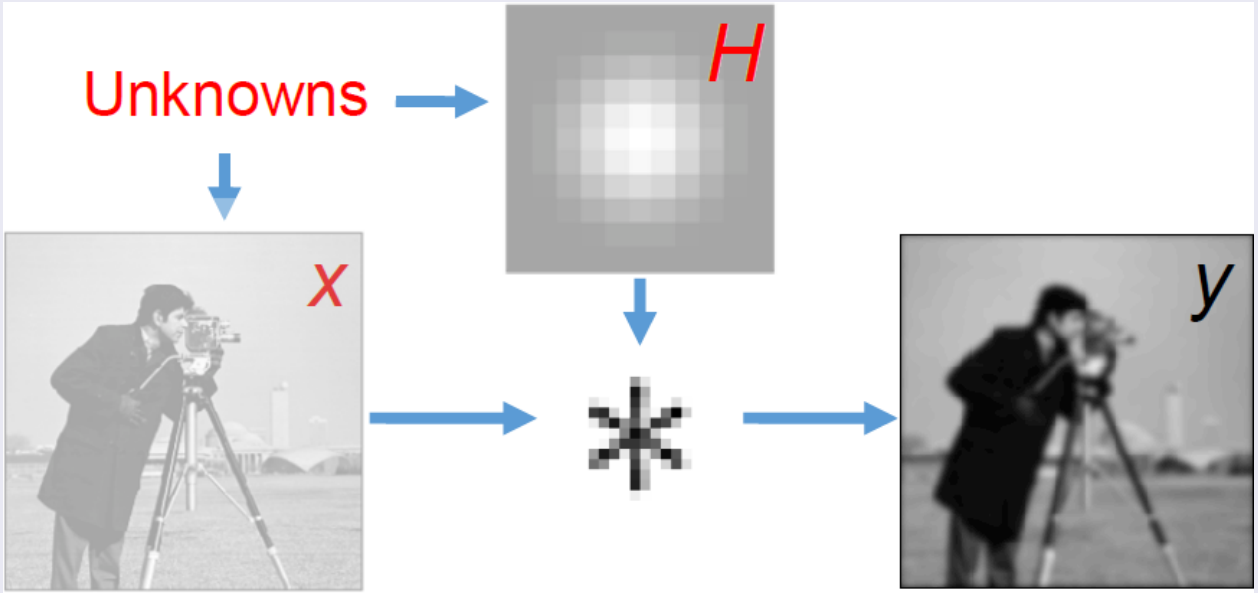


Grey Scale Histogram: LW (left) and IST (right)



Application: Bilinear inverse problem - blind deconvolution

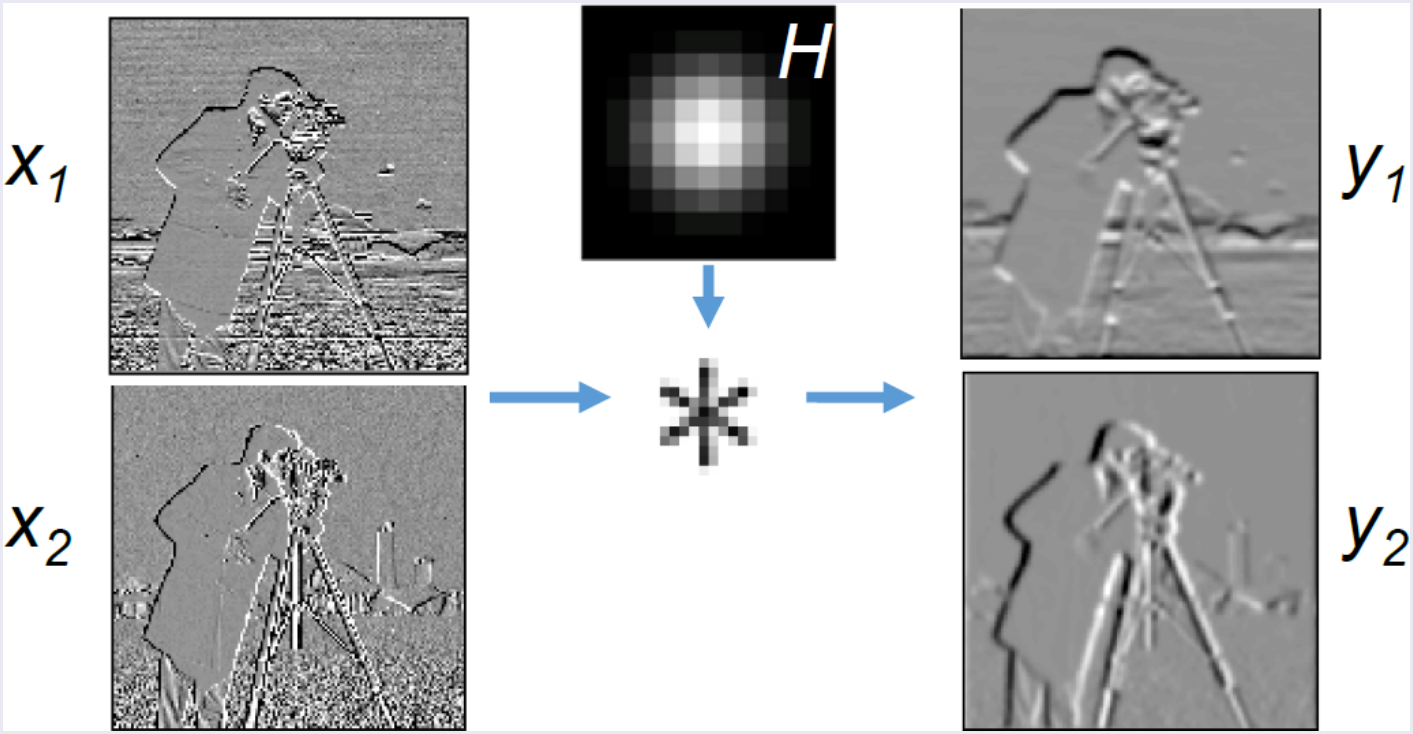
Blind deconvolution problem



- Both $\mathbf{x}$ and $\mathbf{H}$ are unknown
- Can one recover both simultaneously?
- With sufficient constraints $\min_{\mathbf{x},\mathbf{H}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$ is well posed
- First must identify suitable basis on which image is sparse

Blind image deconvolution [Shearer, Gilbert, Hero 13])

Edge decomposition gives sparse coefficients



$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{H}\mathbf{x}_1 \\ \mathbf{H}\mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \end{bmatrix} \quad \Leftrightarrow \quad \mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}$$

Alternating projective gradient [Shearer, Gilbert, H 13]

Consider the minimization problem

$$\min_{\mathbf{X},\mathbf{H}} \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_2^2, \quad \underbrace{\|\mathbf{X}\|_1 \leq \delta}_{\ell_1 \text{ ball}}, \quad \underbrace{\mathbf{H} \geq 0, \sum_{ij} [\mathbf{H}]_{ij} = 1}_{\text{simplex set}}$$

Alternating projective gradient minimization

Define $F(\mathbf{X}, \mathbf{H}) = \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_2^2$.

$$\hat{\mathbf{X}}^{(t+1)} = \text{shrink}_\lambda \left(\hat{\mathbf{X}}^{(t)} + \lambda \nabla_{X^{(t)}} F(\mathbf{X}^{(t)}, \mathbf{H}^{(t)}) \right)$$

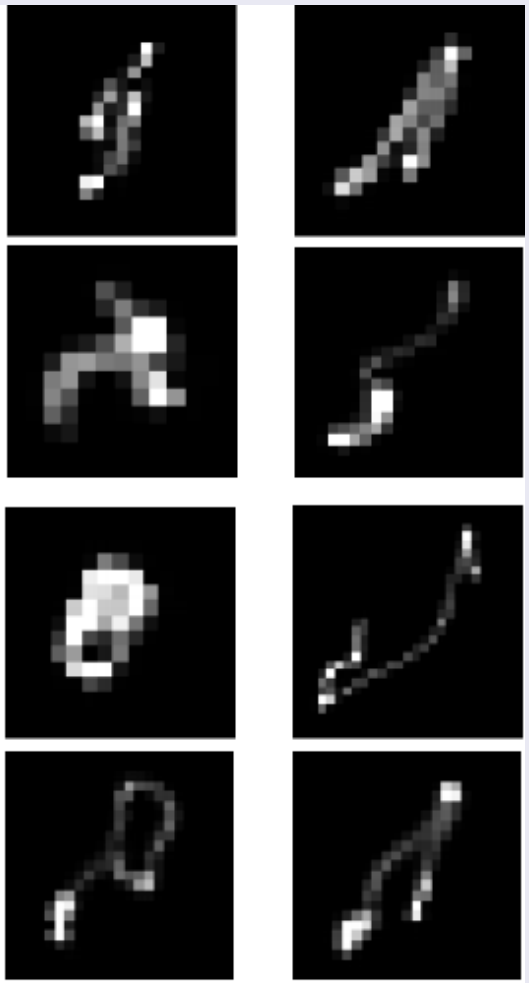
$$\hat{\mathbf{H}}^{(t+1)} = \text{simp}_\lambda \left(\hat{\mathbf{H}}^{(t)} + \lambda \nabla_{H^{(t)}} F(\mathbf{X}^{(t)}, \mathbf{H}^{(t)}) \right)$$

shrink_λ and simp_λ are respective proximity operators

Application: camera jitter [Shearer, Gilbert, H 13])

Levin's corpus of jitter photographs [Levin 2009]

images



Blur kernels

Application: camera jitter [Shearer, Gilbert, H 13])

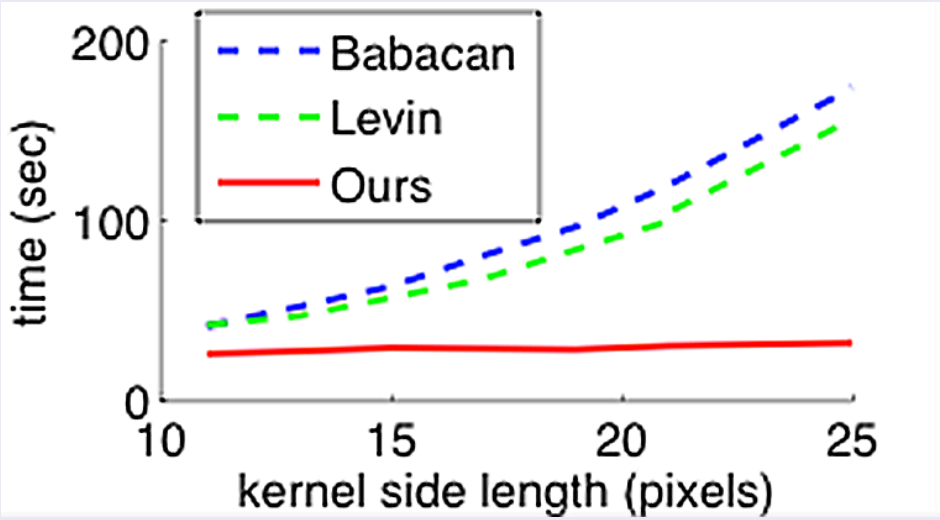
Alternating projected gradient iterations

- PSF and deblurred image are recovered at iteration $t = 150$
- Additional acceleration achieved by multiscale seeding [Krishnan, . . . , 2009]
- Start with downsampled $\mathbf{y}, \mathbf{H}, \mathbf{x}$ and gradually increase resolution over iterations

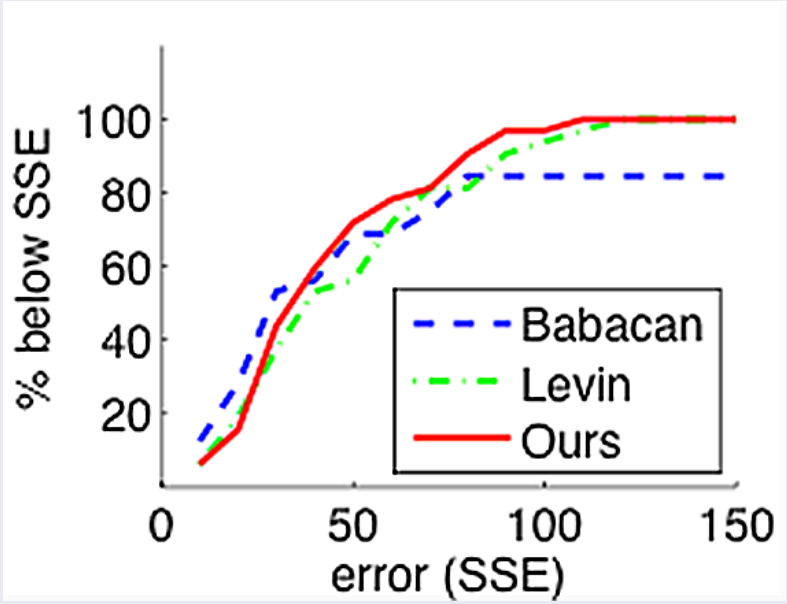
Performance comparisons [Shearer, Gilbert, H 13])

- Compare to state-of-the-art methods [Levin 2009] and [Barbacan 2012] that:
  - Solve inner H-step exactly via quadratic programming
  - Use heavy-tailed Bayesian edge priors on $\mathbf{x}$

Computation speed

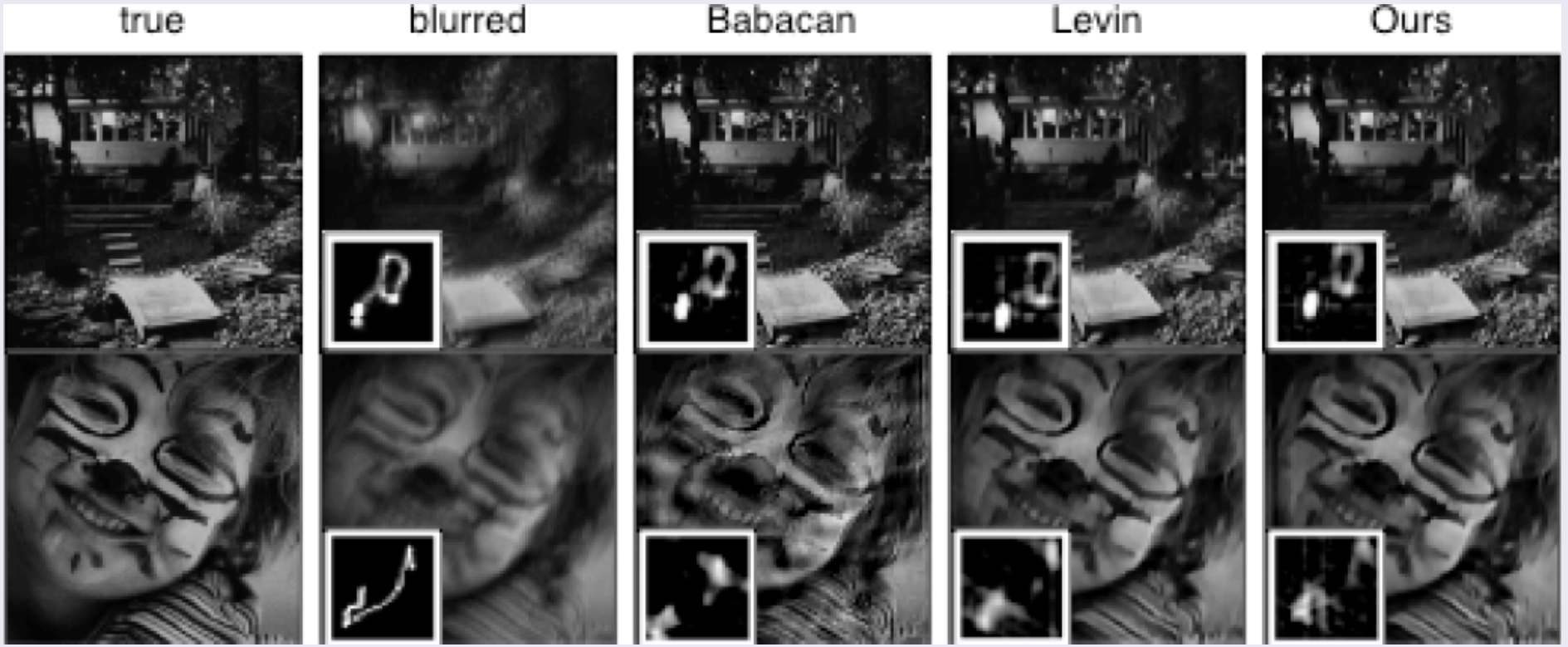


Performance



Visual comparisons [Shearer, Gilbert, H 13])

Visual dejitter comparisons



Alternative optimization perspective: variable splitting

Original IST can be sped-up

- Two-step IST (TwIST) [Boucas-Dias, Figueriedo 07]
- Fast IST (FISTA) [Beck, Teboulle 09]
- SPaRSA [Wright, Nowak, Figueriedo 08]

However, splitting methods can provide additional improvements

- Alternating directions method of multipliers (ADMM) [Growinski, Morrocco 75][Gabay 83], [Ekstein, Bertsekas 92]
- Split-Bregman iterations [Goldstein, Osher 08]
- Optimization transfer [Ortega, Rheinbolt 70], [Lange 00]

Many engineering, CS and statistics applications

- Image restoration and reconstruction [Alfonso, Boucas-Dias, Figueriedo 11], [Ramani Fessler 11]
- Signal processing and vision [Combettes, Pesquet 10], [Goldfarb, Ma 10]
- Machine learning [Banerjee . . . 12], [Ng 10], [Liu, . . . , H 12]

Splitting methods: MM and ADMM [Figueriedo 2012]

There are three forms for the original optimization problem

1. Ivanov form: constrain sparsity, optimize fit to data

$$\min_{\|\mathbf{x}\|_1 \leq \delta} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \tag{2}$$

2. Morozov form: constrain fit to data, optimize sparsity

$$\min_{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 \leq \epsilon} \|\mathbf{x}\|_1 \tag{2}$$

3. Tikhonov form: Method of Multipliers (MM)

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

where λ is an undetermined (Lagrange) multiplier.

Properties

- There exist δ, γ, λ such that 1,2,3 are equivalent [Lorenz, 12]
- Can lead to accelerated computation and memory savings [Ramani and Fessler 2012].
- Especially useful for non-quadratic data fit and mixed constraints

Non-smooth optimization: Method of Multipliers (MM)

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{Q}\mathbf{x}\|_1$$

MM implementation [Ramani and Fessler 2012]

Introduce auxiliary variables $\mathbf{u}$, $\mathbf{v}$ and rewrite Tikhonov form in Ivanov form

$$\min_{\mathbf{x},\mathbf{u},\mathbf{v}} \|\mathbf{y} - \mathbf{u}\|_2^2 + \lambda \|\mathbf{v}\|_1, \quad \mathbf{v} = \mathbf{Q}\mathbf{x}, \ \mathbf{u} = \mathbf{H}\mathbf{x}$$

Defining $g(\mathbf{u}, \mathbf{v}) = \|\mathbf{y} - \mathbf{u}\|_2^2 + \lambda \|\mathbf{v}\|_1$ and $\mathbf{z} = [\mathbf{u}, \mathbf{v}]$, equivalent problem:

$$\min_{\mathbf{x},\mathbf{z}} \ g(\mathbf{z}), \quad \mathbf{z} = \mathbf{C}\mathbf{x}, \qquad \mathbf{C} = \begin{bmatrix} \mathbf{H} \\ \mathbf{Q} \end{bmatrix}$$

Convert back to Tikhonov form using MM

$$\min_{\mathbf{x},\mathbf{z}} \ g(\mathbf{z}) + \boldsymbol{\gamma}^T(\mathbf{z} - \mathbf{C}\mathbf{x}) + \frac{\mu}{2}\|\mathbf{z} - \mathbf{C}\mathbf{x}\|_2^2$$

Non-smooth optimization: Alternating Directions MM

$$\min_{\mathbf{x},\mathbf{z}} \; g(\mathbf{z}) + \boldsymbol{\gamma}^T(\mathbf{z} - \mathbf{C}\mathbf{x}) + \frac{\mu}{2}\|\mathbf{z} - \mathbf{C}\mathbf{x}\|_2^2$$

Alternating Directions Method of Multipliers (ADMM)

Complete square

$$\min_{\mathbf{x},\mathbf{z}} \; g(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{z} - \boldsymbol{\eta} - \mathbf{C}\mathbf{x}\|_2^2 \tag{3}$$

where $\boldsymbol{\eta} = -\frac{1}{\mu}\boldsymbol{\gamma}$.
ADMM implements (3) by block coordinate descent over each of $\mathbf{x}$ and $\mathbf{z}$

$$\begin{aligned}
\mathbf{x}^{(t+1)} &= \text{argmin}_{\mathbf{x}} \left\|\mathbf{z}^{(t)} - \mathbf{C}\mathbf{x} - \boldsymbol{\eta}^{(t)}\right\|_2^2 \\
\mathbf{z}^{(t+1)} &= \text{argmin}_{\mathbf{z}} g(\mathbf{z}) + \left\|\mathbf{z} - \mathbf{C}\mathbf{x}^{(t)} - \boldsymbol{\eta}^{(t)}\right\|_2^2 \\
\boldsymbol{\eta}^{(t+1)} &= \boldsymbol{\eta}^{(t)} - (\mathbf{z}^{(t+1)} - \mathbf{C}\mathbf{x}^{(t)})
\end{aligned}$$

Non-smooth optimization: ADMM

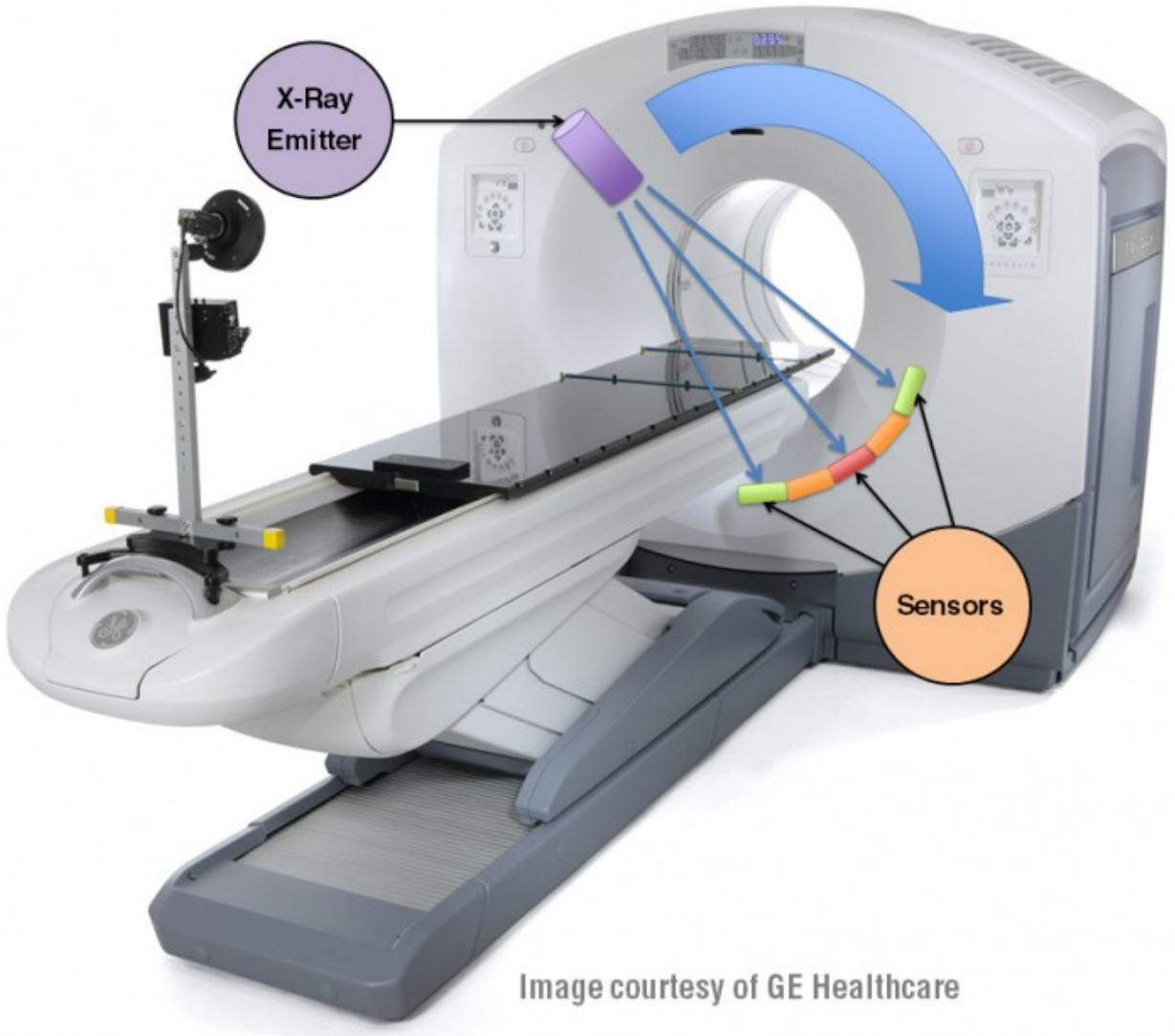
High level view: ADMM alternates between

$$\min_{\mathbf{x}} \; g(\mathbf{z}^{(t)}) + \frac{\mu}{2}\|\mathbf{z}^{(t)} - \boldsymbol{\eta}^{(t)} - \mathbf{C}\mathbf{x}\|_2^2, \quad \min_{\mathbf{z}} \; g(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{z} - \boldsymbol{\eta}^{(t)} - \mathbf{C}\mathbf{x}^{(t)}\|_2^2$$

Properties of ADMM

- General ADMM approach introduced by [Glowinski and Marrocco 1975].
- Can significantly outperform IST and its variants [Figueriedo 2012]
- ADMM will converge under conditions on μ, $\mathbf{H}$ and $\mathbf{Q}$ [Eckstein and Bertsekas, 1992]
- Convergence holds even if minimizations are approximate
- ADMM decouples the integrated problem into easier sub-problems

Illustration: low dose X-ray CT reconstruction



Advantage of iterative statistical reconstruction [Fessler 2012]

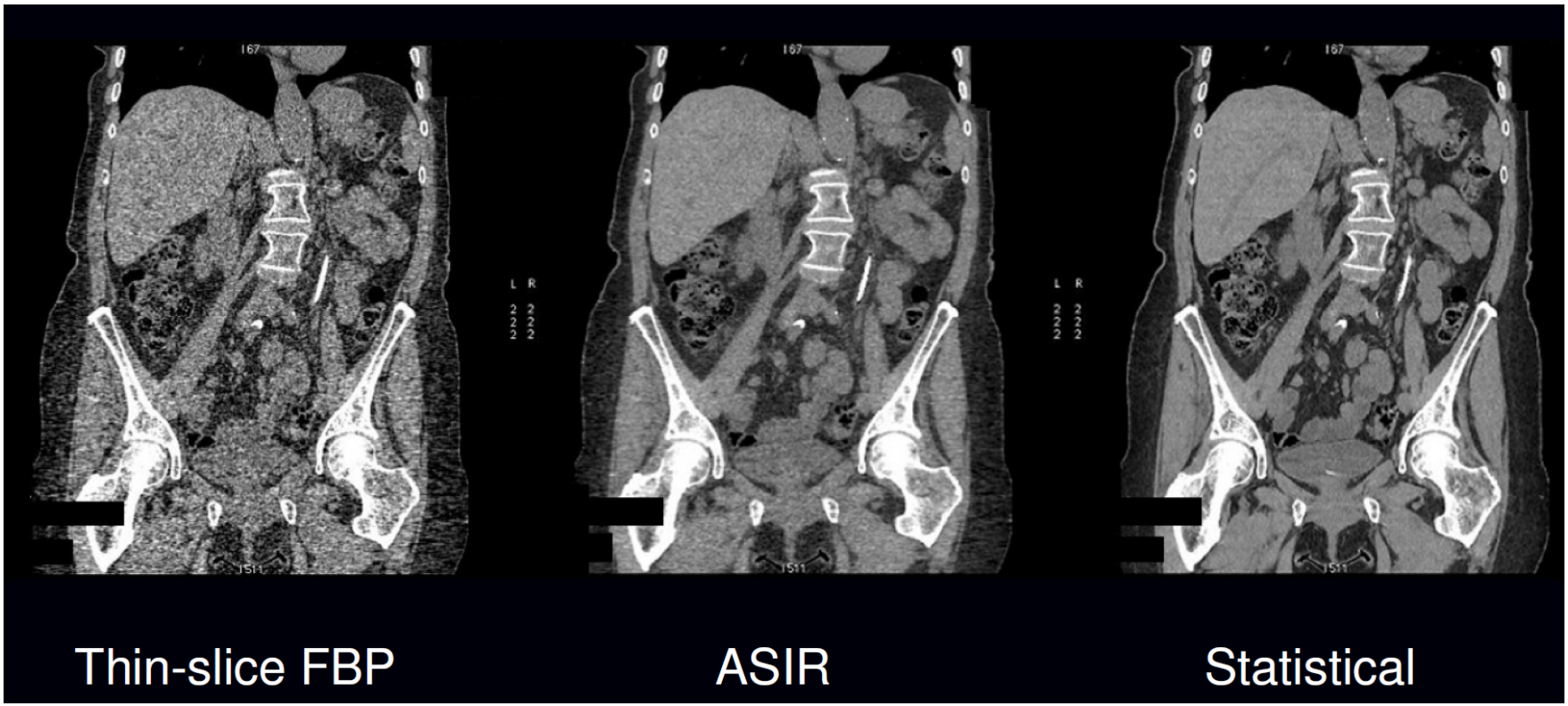


Illustration: X-ray CT reconstruction [Ramani and Fessler 2012]

Reconstruction

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_W^2 + \lambda \|\mathbf{Q}\mathbf{x}\|_1$$

- $\|\mathbf{z}\|_W = \mathbf{z}^T \mathbf{W} \mathbf{z}$ where $\mathbf{W} = \mathrm{diag}(\mathbf{y})$ (Poisson statistics)
- $\mathbf{Q}$ is Haar wavelet basis
- $\mathbf{x}$: $512 \times 513 \times 800 \approx 2 \cdot 10^8$ unknown image parameters
- $\mathbf{y}$: $888 \times 64 \times 7000 \approx 4 \cdot 10^8$ projections (sinogram)
- Projection $\mathbf{H}\mathbf{x}$ and back-projection $\mathbf{H}^T\mathbf{y}$ dominate computations
- $\mathbf{H}$ is sparse matrix corresponding to GE X-ray scanner
- Cone preconditioner applied to $\mathbf{H}$ to make $\mathbf{H}^T\mathbf{H}$ closer to diagonal
- Ground truth $\mathbf{x}$ was obtained by 5000 iterations of NCG

Illustration: low dose X-ray CT reconstruction

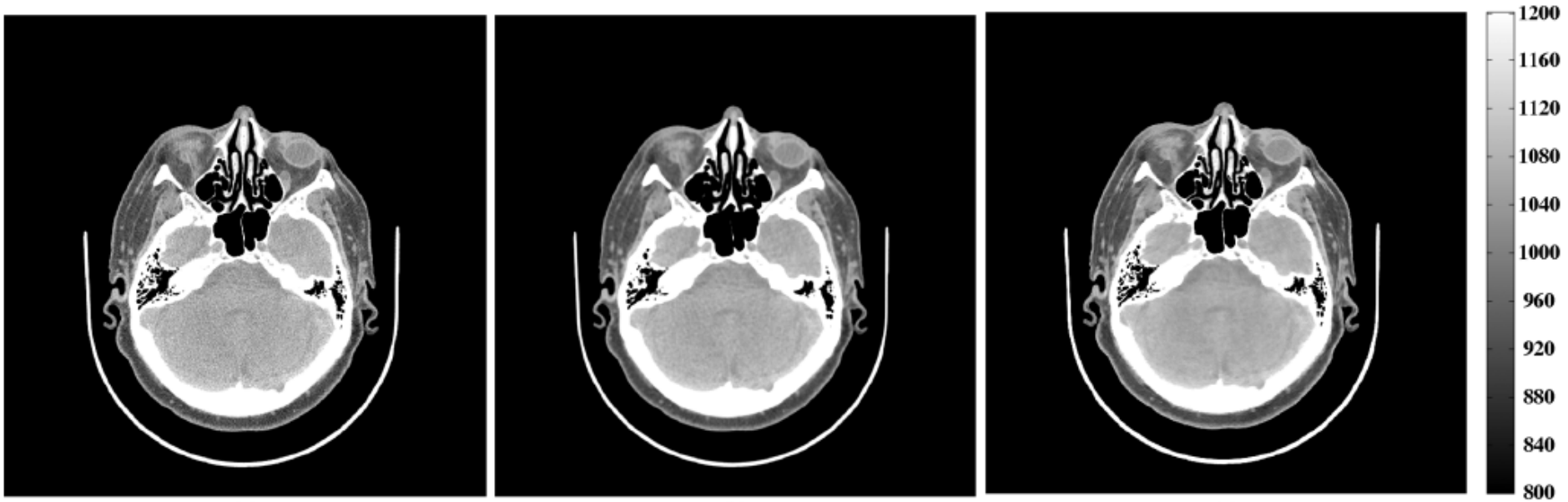


Fig. 4. Experiment with the *in vivo* human head data-set: (a) FBP reconstruction with the ramp filter, also the initial guess $\mathbf{x}^{(0)}$ for all iterative algorithms, (b) RBP reconstruction with Hanning filter, and (c) PWLS reconstruction with the strictly convex regularizer (23), also the unique solution $\mathbf{x}^\star$ to $\mathbf{P0}$. Images in (a)–(c) are displayed in Hounsfield units indicated beside (c). The regularized reconstruction (c) is less noisy and preserves anatomical features compared to both FBP results.

Illustration: low dose X-ray CT reconstruction

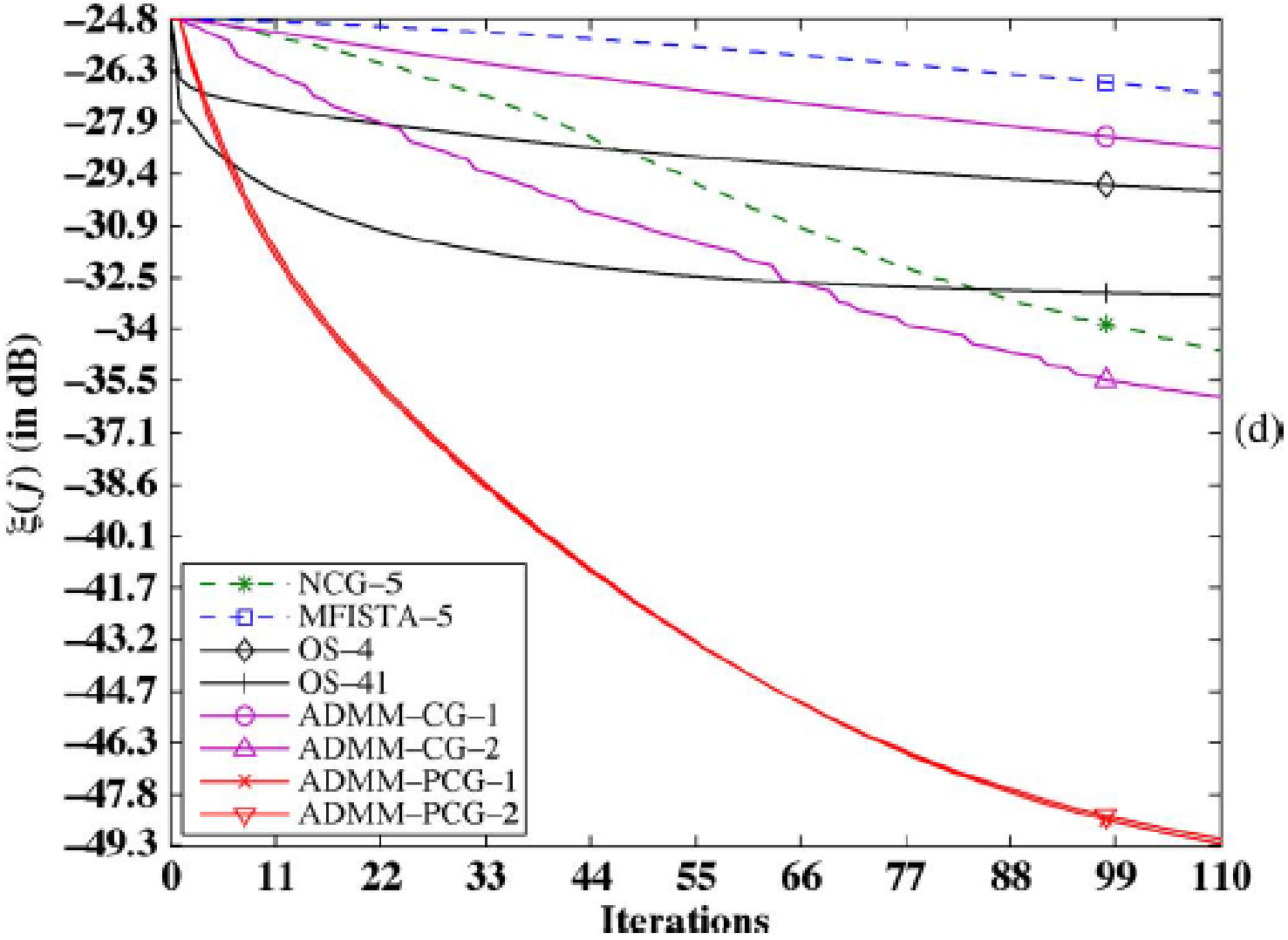
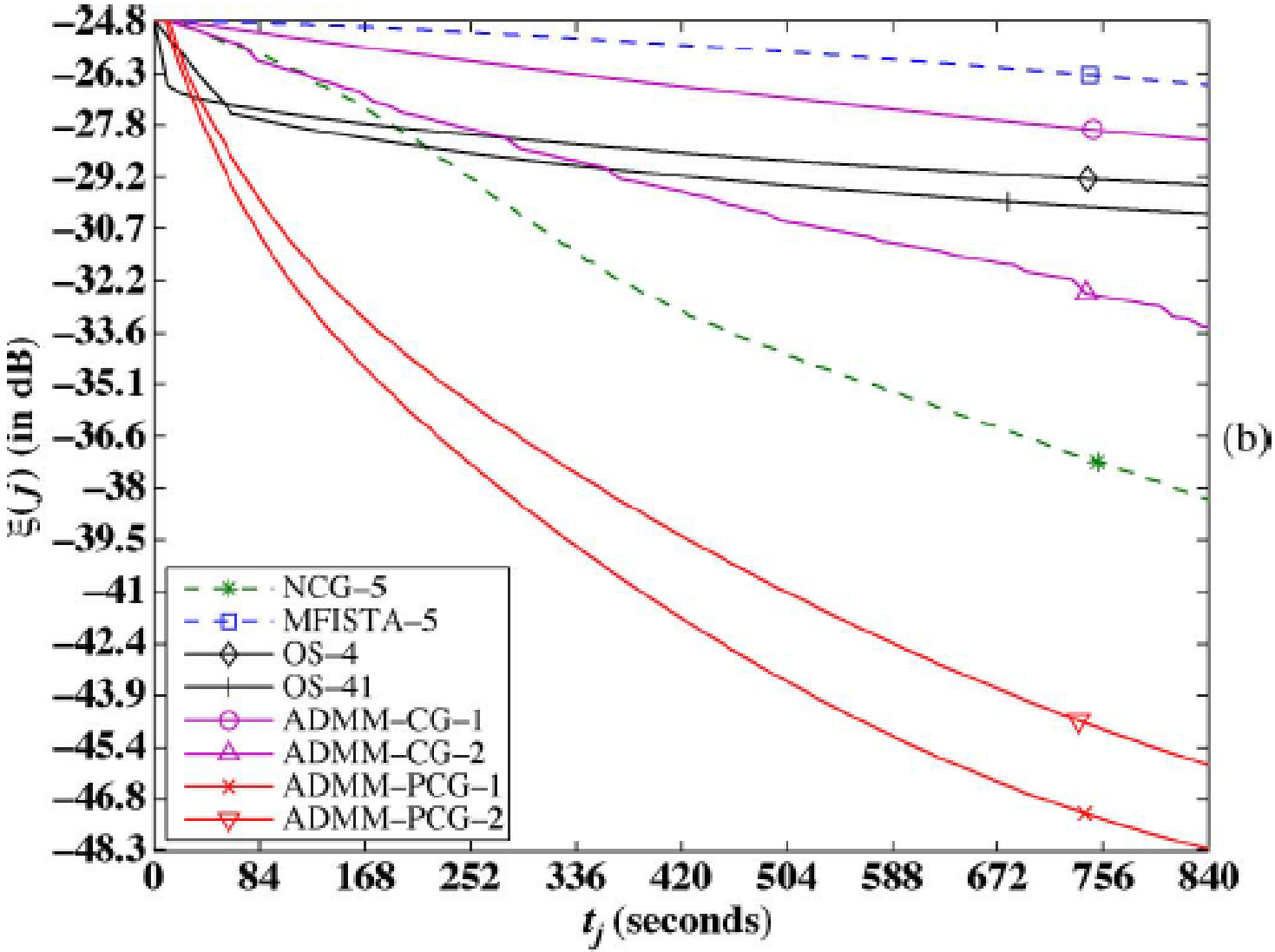


Illustration: low dose X-ray CT reconstruction



4. Bayesian sparse reconstruction methods

Observation model is again

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \quad \mathbf{x} \in \mathcal{C}$$

Bayesian approach yields posterior distribution $\Leftrightarrow$ inverse probability

Define:

- $\mathbf{v}$: a random noise process with prior density $f_v(\mathbf{v})$
- $\mathbf{x}$: a random signal process with prior density $f_x(\mathbf{x})$

Obtain:

$$f_{y|x}(\mathbf{y}|\mathbf{x}) = f_v(\mathbf{y} - \mathbf{H}\mathbf{x}), \qquad \text{(Likelihood function)}$$

$$f_{x|y}(\mathbf{x}|\mathbf{y}) = \frac{f_{y|x}(\mathbf{y}|\mathbf{x}) f_x(\mathbf{x})}{\int_X f_{y|x}(\mathbf{y}|\tilde{\mathbf{x}}) f_x(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}}, \qquad \text{(Posterior density)}$$

These characterize forward model and inverse model, respectively.

Bayesian reconstruction methods

From posterior density $f_{x|y}$ can extract much information about $\mathbf{x}$

- minimum mean squared error (MSE) reconstruction:

$$\hat{\mathbf{x}}_{MSE} = E[\mathbf{x}|\mathbf{y}] = \int_X \mathbf{x} f_{x|y}(\mathbf{x}|\mathbf{y}) d\mathbf{x}$$

- MAP reconstruction

$$\hat{\mathbf{x}}_{MAP} = \text{argmax}_{\mathbf{x}} f_{x|y}(\mathbf{x}|\mathbf{y})$$

- Posterior MSE

$$\hat{\sigma}^2 = E[\|\mathbf{x} - \hat{\mathbf{x}}_{MSE}\|_2^2|\mathbf{y}]$$

- Posterior confidence intervals and regions

$$P(\mathbf{x} \in \mathcal{R}|\mathbf{y}) = \int_{\mathcal{R}} f_{x|y}(\mathbf{x}|\mathbf{y}) d\mathbf{x}$$

- No tuning parameters (k, δ, ϵ, λ, μ) once f_x, $f_{y|x}$ are specified

IST as special case of Bayesian reconstruction

Non-smooth optimization as MAP estimation

- Reconsider Tikhonov's form of ℓ_1-sparsity constrained optimization

$$\hat{\mathbf{x}} = \text{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

- Bayesian MAP reconstruction

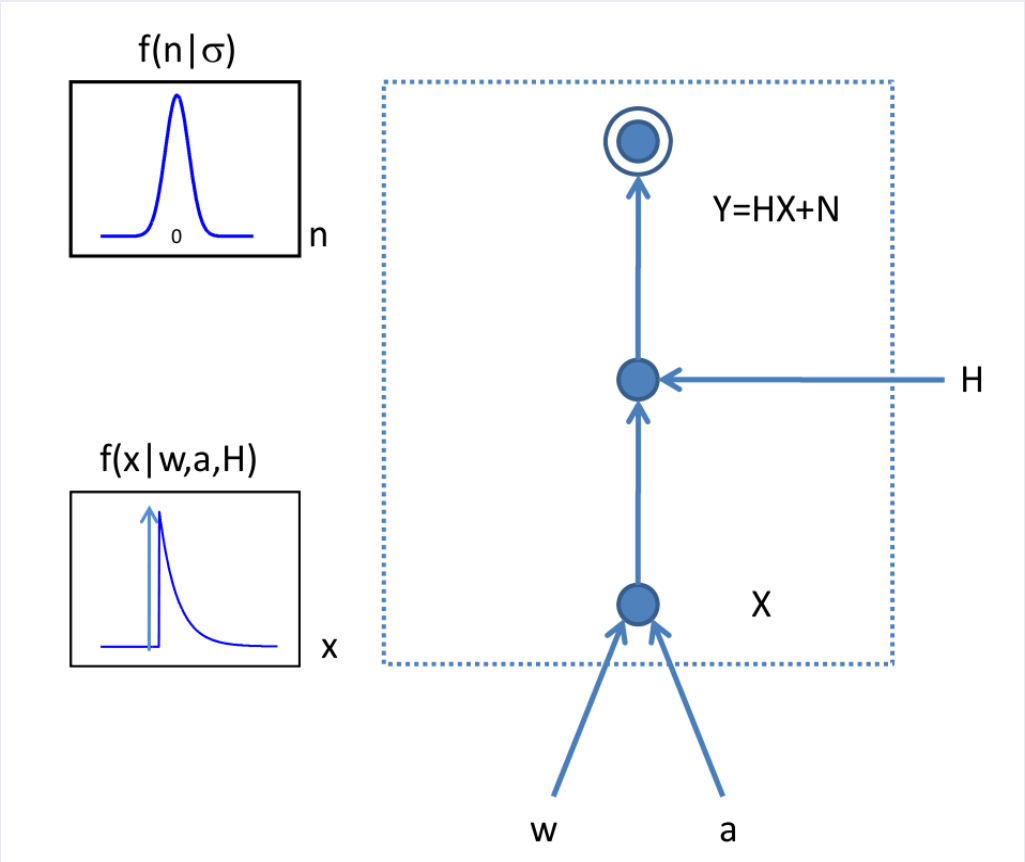
$$\hat{\mathbf{x}} = \text{argmax}_{\mathbf{x}} \log f(\mathbf{x}|\mathbf{y}) = \text{argmax}_{\mathbf{x}} \ \log f_v(\mathbf{y} - \mathbf{H}\mathbf{x}) + \log f_x(\mathbf{x})$$

- Assume f_v is Gaussian white noise and f_x is i.i.d Laplacian

$$\begin{aligned} f_v(\mathbf{y} - \mathbf{H}\mathbf{x}) &= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{H}\mathbf{x})^T(\mathbf{y} - \mathbf{H}\mathbf{x})\right) \\ f_x(\mathbf{x}) &= \frac{1}{(2a)^n} \exp\left(-\frac{1}{a}\|\mathbf{x}\|_1\right) \end{aligned}$$

Graphical model: Laplace+zero (LAZE) prior

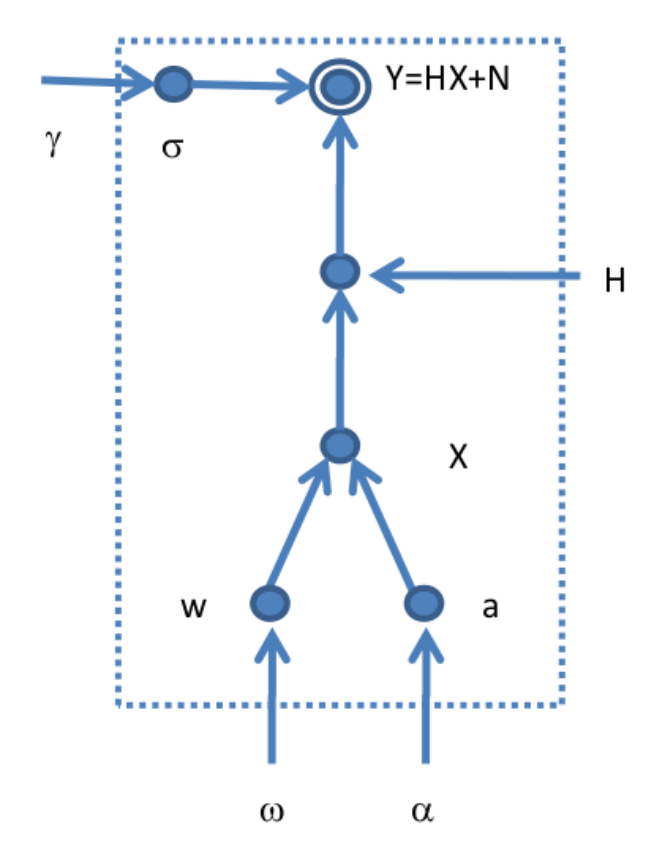
Synthesis: image and noise priors



$$f(\mathbf{y}|\mathbf{x}, w, a, H) = f_N(\mathbf{y} - \boldsymbol{H}\mathbf{x}|\sigma)$$

Hierarchical graphical model: hyperparameters γ, ω, α

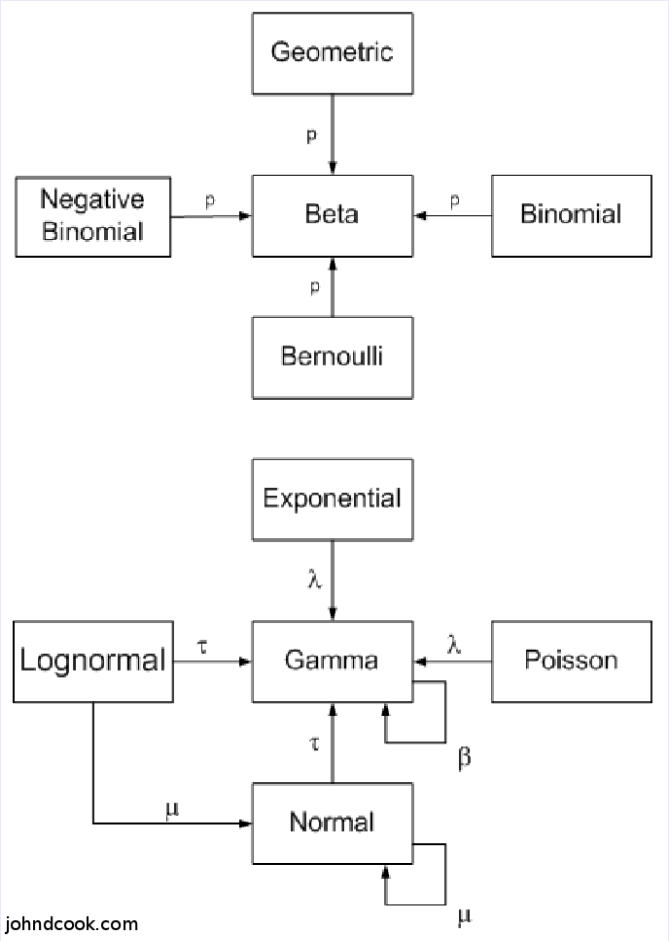
Hierarchical Bayesian graphical model



$$f(\mathbf{y}, \mathbf{x}|H) = \int f(\mathbf{y}, \mathbf{x}|w, a, H) f(w, a|\omega, \alpha) dwda$$

Constellation of Bayesian conjugate priors

Conjugate priors and their relations



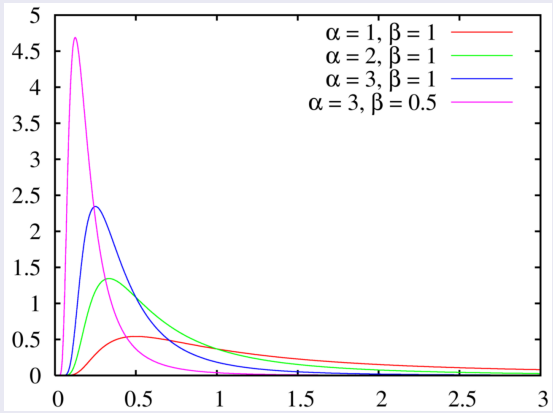
$$f(\mathbf{y}, \mathbf{x}|H) = \int f(\mathbf{y}, \mathbf{x}|w, a, H) f(w, a|\omega, \alpha) dwda$$

Important to make judicious choice of priors

Conjugate priors

- Beta sparsity prior: $w|\boldsymbol{\omega} \sim \mathcal{B}(\omega_1, \omega_0)$,
- Inverse-gamma mean-intensity: $a|\boldsymbol{\alpha} \sim \mathcal{IG}(\alpha_0, \alpha_1)$,
- Inverse-gamma variance prior: $\sigma|c, \kappa \sim \mathcal{IG}(c; 1, \kappa/2)$

Inverse gamma prior



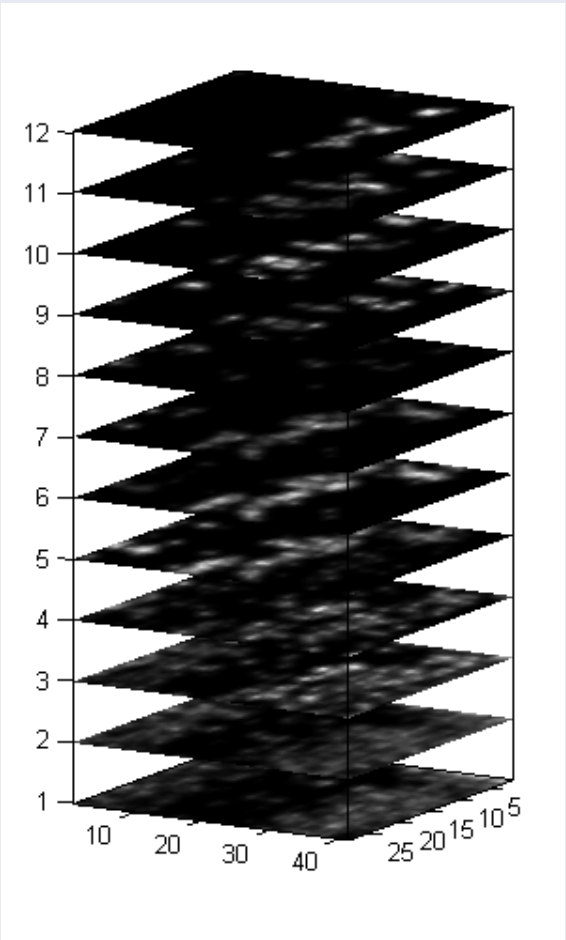
Full posterior distribution $f(\mathbf{x}|\mathbf{y})$ is closed form

$$f(\mathbf{x}|\mathbf{y}) \propto \frac{B(\omega_1 + n_1, \omega_0 + n_0)}{\|\mathbf{y} - T(\boldsymbol{\kappa}, \mathbf{x})\|^m} \frac{\Gamma(n_1 + \alpha_0)}{[\|\mathbf{x}\|_1 + \alpha_1]^{n_1 + \alpha_0}}.$$

where $n_1 = \|\mathbf{x}\|_0$ and $n_0 = n - \|\mathbf{x}\|_0$.

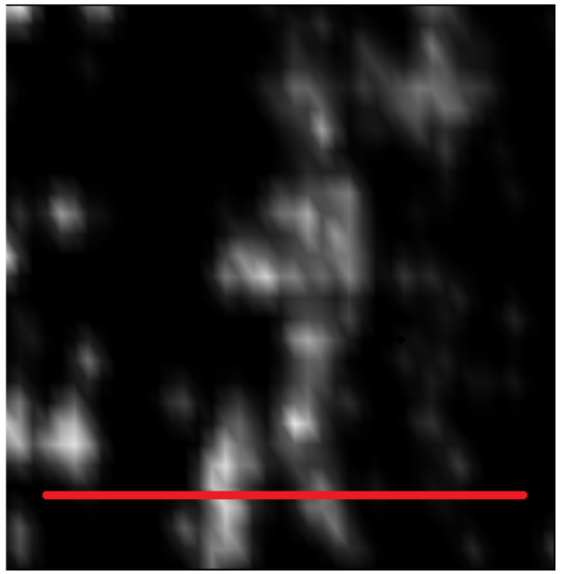
MRFM: Bayesian reconstruction of tobacco virus data

3D view of reconstruction

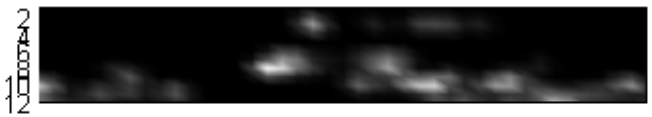


$\Delta z = 4.3nm$, pixel spacing is $8.3nm \times 16.6nm$ in $x \times y$, respectively. The size of (x, y) plane is $498nm \times 531.2nm$.

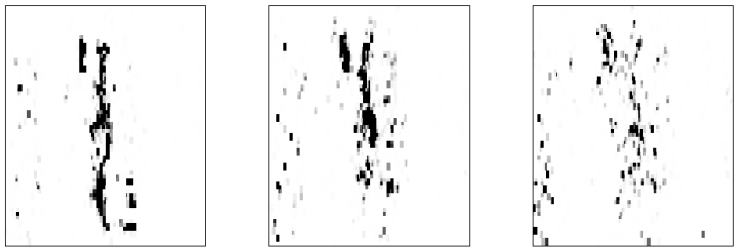
Views on x-y plane

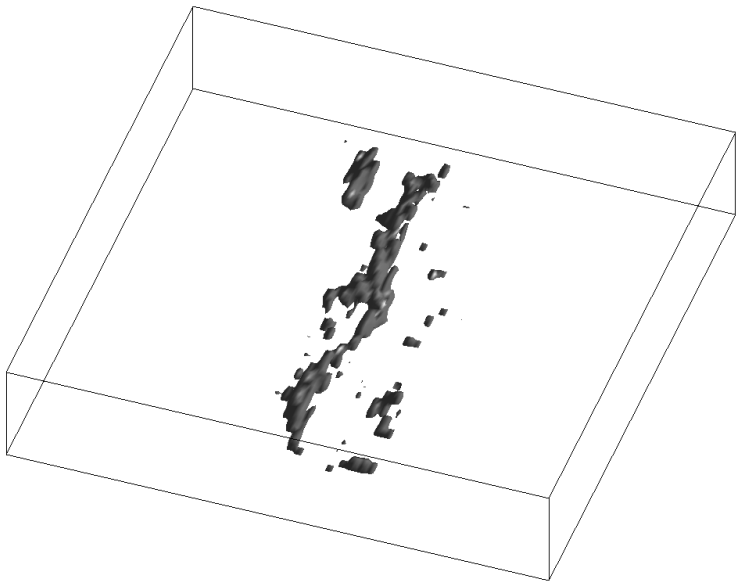


x-z plane (cut in red line)



Reconstruction





HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

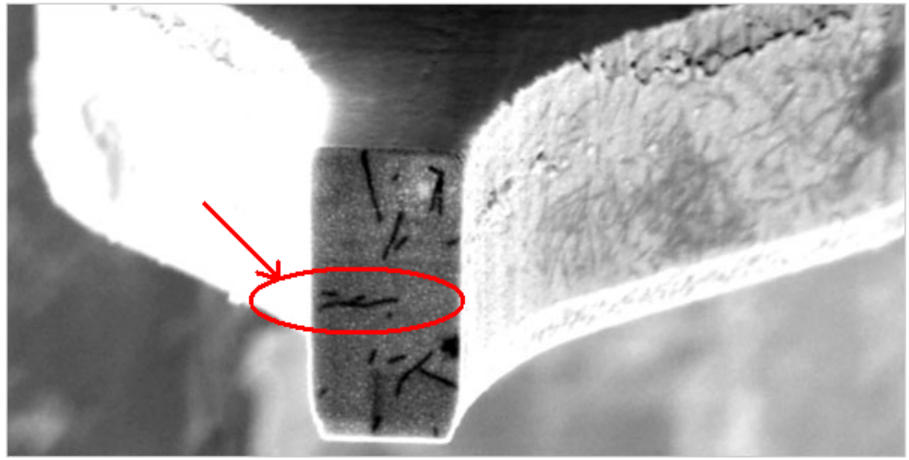
The New York Times

Science

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

ENVIRONMENT | SPACE & COSMOS

A Breakthrough in Imaging: A New Way to See a Virus



The end of the cantilever arm, with virus particles attached.

By JOHN MARKOFF
Published: January 12, 2009

Researchers at an I.B.M. laboratory have captured a three-dimensional image of a biological virus using, for the first time, a technique that has some similarity to magnetic resonance imaging, a tool routinely used by physicians to peer inside the human body.

SIGN IN TO E-MAIL OR SAVE THIS
PRINT
REPRINTS
SHARE
SHARE

Related

Web Link

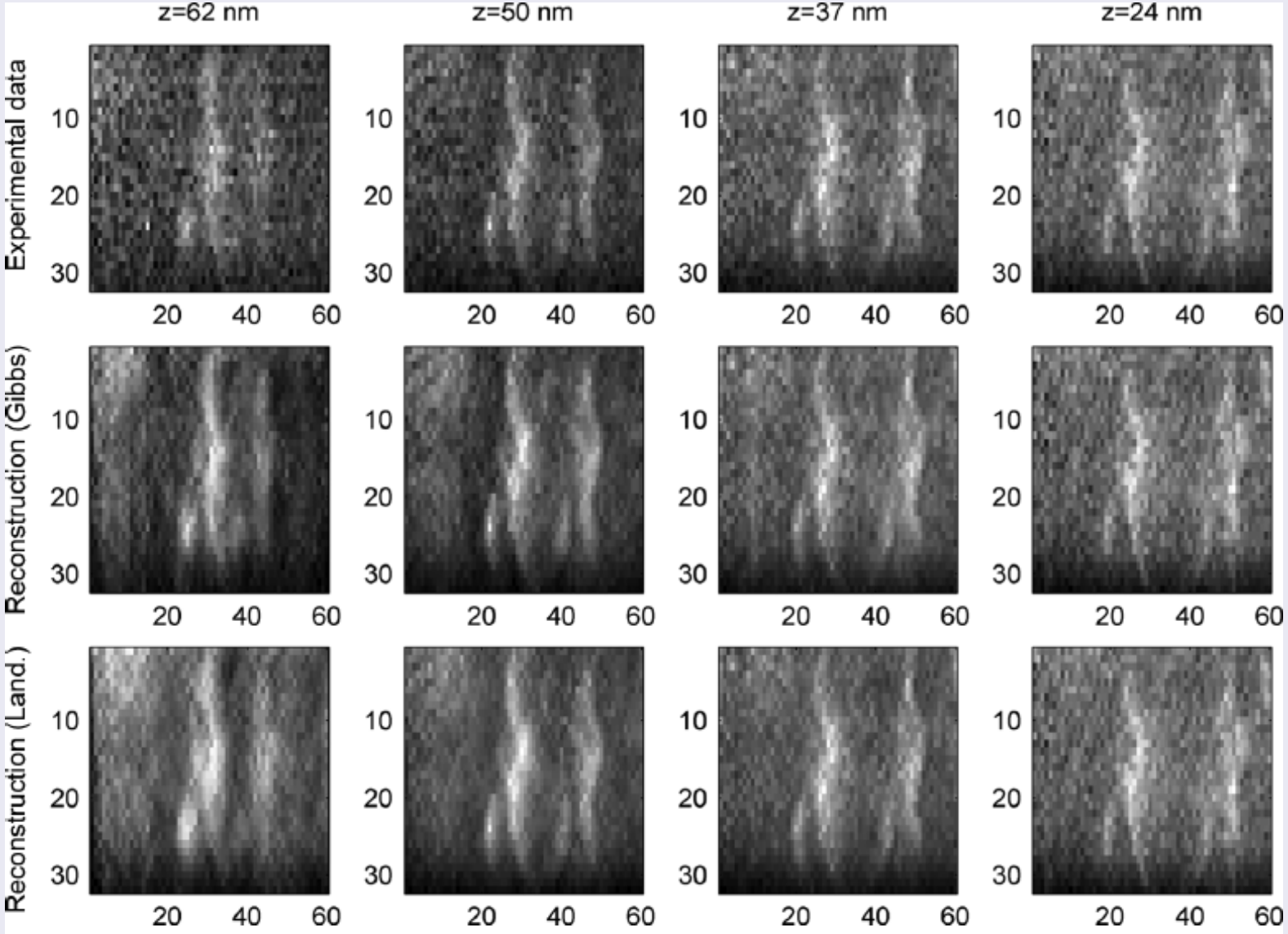
Nanoscale Magnetic Resonance Imaging (PNAS)

RSS Feed

Although the technique is akin to M.R.I., the results were 100 million times better in terms of resolution with the new technique, magnetic resonance force microscopy, or M.R.F.M. The team of researchers, based at the computer maker's Almaden Research Center

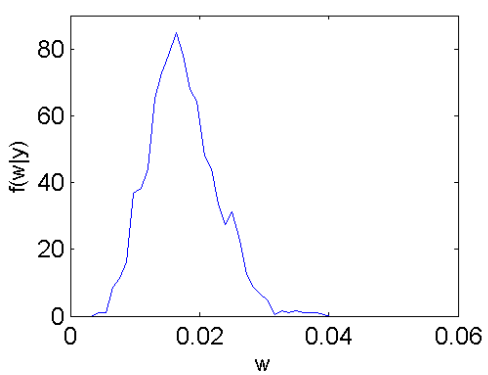
Visual comparisons for Tobacco virus study

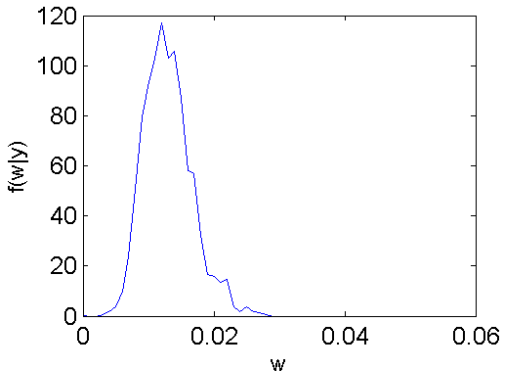
Observed vs Bayesian vs Landweber reprojections



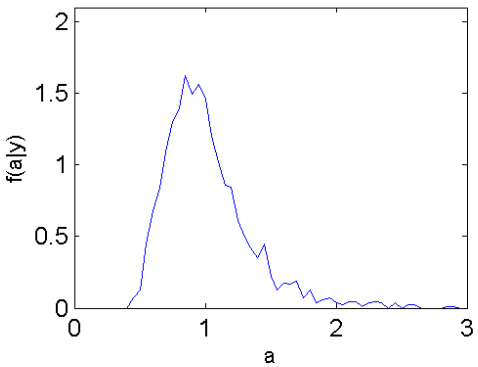
Posterior distributions [Dobigeon, Tourneret, H 09]

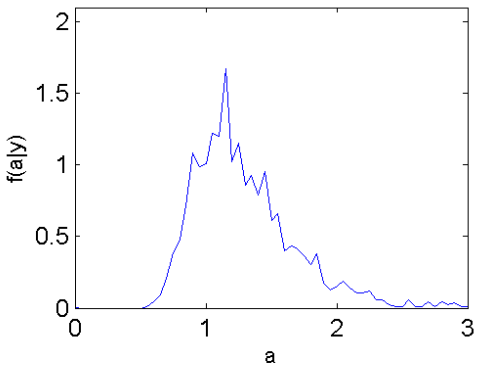
Posterior distribution of w





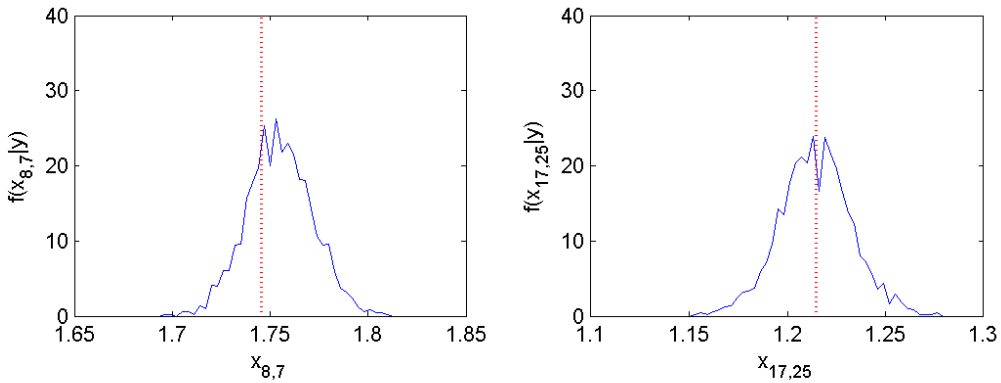
Posterior distribution of a



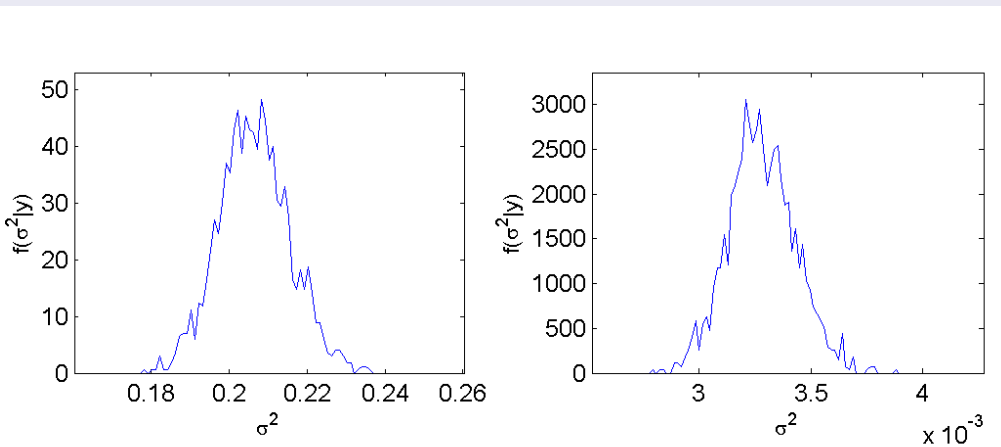


Posterior distributions [Dobigeon, Tourneret, H 09]

Posterior distribution of non-zero pixels



Posterior distribution of σ^2



5. Conclusions

- Sparse constraints are ubiquitous in imaging
- The field of sparse reconstruction has considerably matured over the past 10 years
  - ADMM gives unprecedented accuracy and speed improvements
  - Bayesian methods are slower but bring uncertainty quantification and increased stability (no tuning)
- Other relevant areas not covered here
  - Markov spatial dependency: [Park, Dobigeon, H, 13]
  - Semi-blind to PSF: [Park,Dobigeon, H 12&13]
  - Dictionary-based methods: [Park,. . . , H 13], [Chen, . . . , H 14]
  - Covariance inverse problems: [H, Rajaratnam 11&12], [Firouzi, Rajaratnam, H 13]
  - Streaming reconstruction (spatio-temporal): [Chen, H 11]
  - Applications to materials science: [Park,. . . , H 2013]