# Estimation, Detection and Filtering of Medical Images

Ashish Raj, PhD

Co-Director, Image Data Evaluation and Analytics Laboratory (IDEAL) Department of Radiology Weill Cornell Medical College New York

### Outline

- Part I : Basics of Medical image Filtering and Convolution
- Part II : Estimation Theory and examples
- Part III: Detection Theory and examples

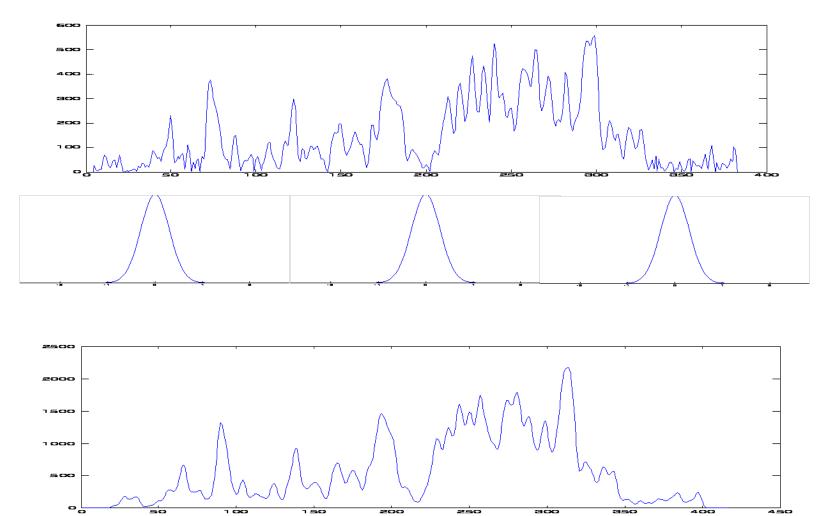
### **Convolution Basics**

Convolution is defined as

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} y(\tau)h(t-\tau)d\tau$$

- Practically achieved as follows:
  - Flip h(t)
  - Slide it into x(t) by amount tau
  - At each position tau, calculate the area of overlap between x and h

### Example



зso

50

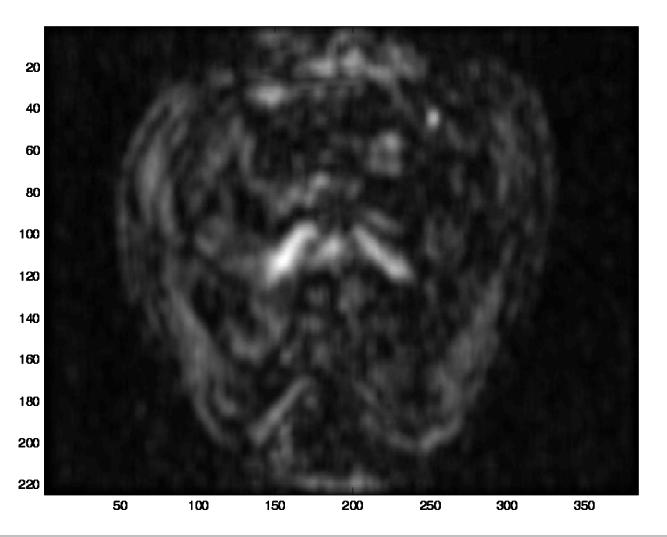
IDEA Lab, Radiology, Cornell

300

400

450

### **Filtering and Convolution**



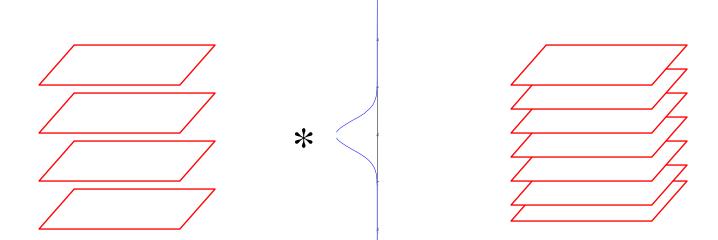


IDEA Lab, Radiology, Cornell

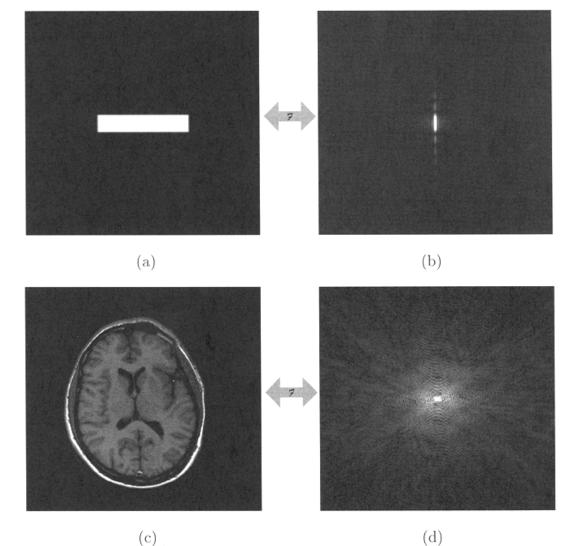
## Convolution For Interpolation and Resampling

- Sometimes need to "fill in" missing data
- Interpolation to resample image on finer grid
- Resampling is used to change the "nonminal" resolution of images
- Example: if multi-slice images with non-isotropic resolution, resampling can make it isotropic
- IMPORTANT: resampling, filtering or interpolation does NOT increase "actual" resolution

### **Resampling example**



### k-space and image-space



k-space & imagespace are related by the 2D FT

IDEA Lab, Radiology, Cornell

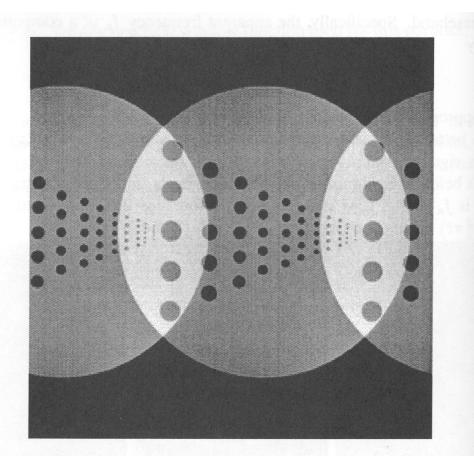
# How many points do we need to sample?

- $\Delta k = 1/FOV$
- Why? Due to the Sampling Theorem

"Suppose a signal I(x) is non-zero only within [-W/2, W/2]. Then its Fourier transform  $F(I)(k_x)$  must be sampled at least as densely as  $\Delta k_x = 1/W$ ."

- Note this works regardless of direction of transform (Duality property)
- What happens if this is violated? ALIASING

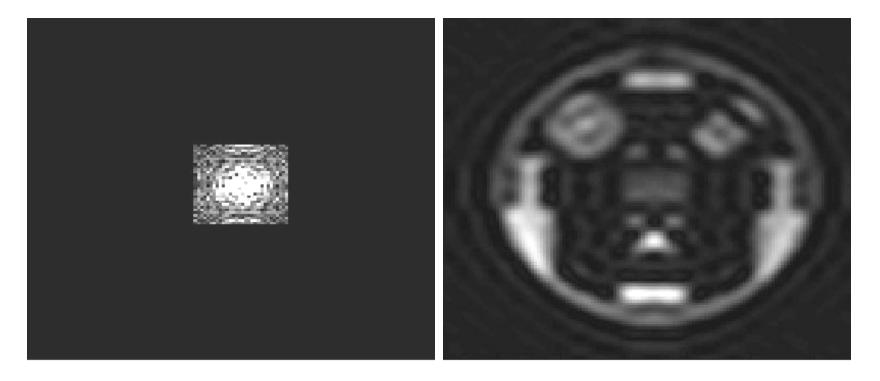
## **Aliasing Example**



**Figure 8.9** Aliasing artifacts due to undersampling along the horizontal direction by a factor of two.

### **Truncation**

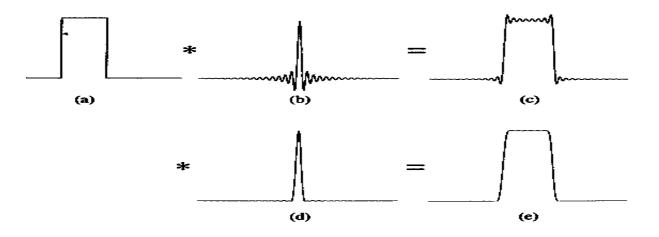
Truncation = sampling central part of k-space



## **Truncation**

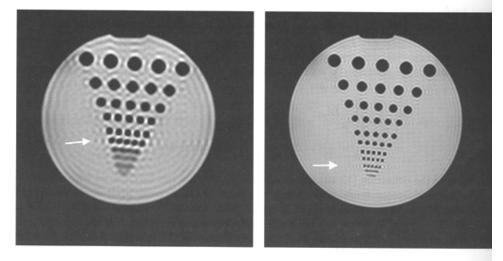
- Both blurring and ringing are a result of truncation
- Intrinsic resolution = size of the blur
- To reduce blur (hence increase resolution) we need to sample up to a larger k-space radius
- Can characterise resolution by the point spread function (PSF) which is simply the blurring kernel
- Note: zero-padding can increase matrix size but can not increase resolution!

# Ringing



- ringing can be reduced by multiplying the signal by a smooth window - called windowing
- Popular window choices:
  - Kaiser-Bessel
  - Hanning
  - Raised cosine

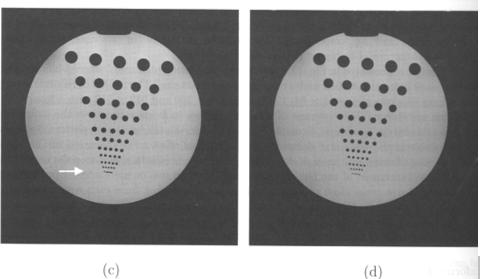
### **Ringing Example**



(a)



(b)



### Part II : Estimation Theory and Examples

- Introduction to optimal estimates
- Different types of optimal estimates
- Estimation examples from MR

# **Estimation Theory**

Consider a linear process

 $y = H \theta + n$ 

- y = observed data
- $\theta$  = set of model parameters
- n = additive noise
- Then Estimation is the problem of finding the statistically optimal θ, given y, H and knowledge of noise properties
- MR is full of estimation problems

## **Different approaches to estimation**

- Minimum variance unbiased estimators
- Least Squares
  Maximum-likelihood
  Maximum entropy
  Maximum a posteriori
  uses knowledge of noise PDF

uses prior information about θ

## **Least Squares Estimator**

• Least Squares:

 $\theta_{LS} = argmin ||y - H\theta||^2$ 

- Natural estimator want solution to match observation
- Does not use any information about n
- There is a simple solution (a.k.a. pseudo-inverse):
   θ<sub>LS</sub> = (H<sup>T</sup>H)<sup>-1</sup> H<sup>T</sup>y

What if we know something about the noise? Say we know Pr(n)...

### Maximum Likelihood Estimator

- Simple idea: want to maximize  $Pr(y|\theta)$
- Can write  $Pr(n) = e^{-L(n)}$ ,  $n = y H\theta$ , and  $Pr(n) = Pr(y|\theta) = e^{-L(y, \theta)}$
- if white Gaussian n,  $Pr(n) = e^{-||n||^2/2\sigma^2}$  and  $L(y, \theta) = ||y-H\theta||^2/2\sigma^2$ 
  - $\theta_{ML} = \operatorname{argmax} \Pr(y|\theta) = \operatorname{argmin} L(y, \theta)$
  - called the likelihood function
    - $\theta_{ML} = argmin ||y-H\theta||^2/2\sigma^2$
- This is the same as Least Squares!

## Maximum Likelihood Estimator

- But if noise is jointly Gaussian with cov. matrix C
- Recall C, E(nn<sup>T</sup>). Then

 $Pr(n) = e^{-\frac{1}{2}n^{T}C^{-1}n}$ 

 $L(y|\theta) = \frac{1}{2} (y-H\theta)^{T} C^{-1} (y-H\theta)$ 

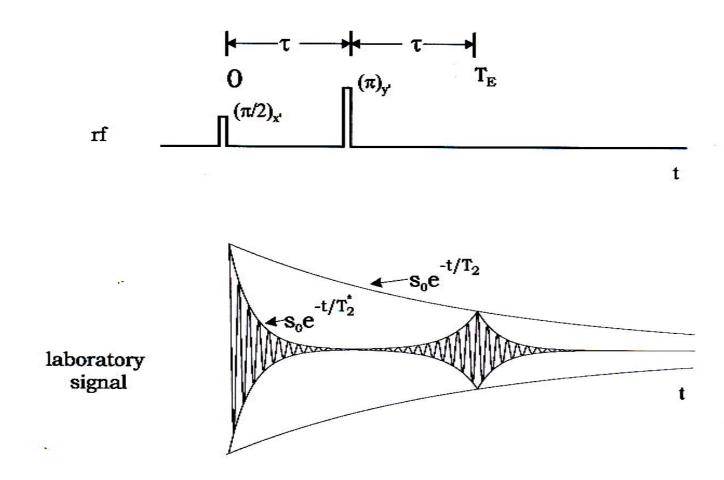
 $\theta_{ML} = \operatorname{argmin} \frac{1}{2} (y-H\theta)^{T}C^{-1}(y-H\theta)$ 

• This also has a closed form solution  $\theta_{ML} = (H^T C^{-1} H)^{-1} H^T C^{-1} y$ 

 If n is not Gaussian at all, ML estimators become complicated and non-linear

Fortunately, in MR noise is usually Gaussian

# Example - estimating T<sub>2</sub> in repeated spin echo data



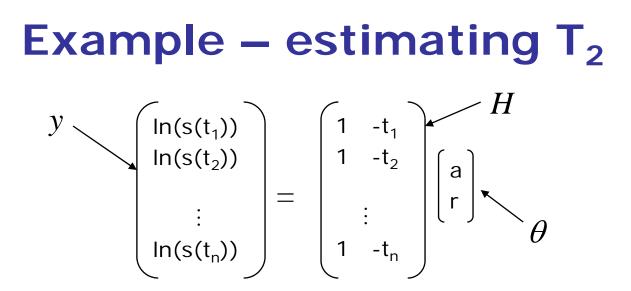
# Example – estimating T<sub>2</sub> in repeated spin echo data

 $s(t) = e^{-t/T_2} \int dr \rho(r)$ 

Need only 2 data points to estimate T<sub>2</sub>:

 $T_{2est} = [T_{E2} - T_{E1}] / ln[s(T_{E1})/s(T_{E2})]$ 

- However, not good due to noise, timing issues
- In practice we have many data samples from various echoes



Least Squares estimate:

$$\theta_{LS} = (H^T H)^{-1} H^T y$$
$$T_2 = 1/r_{LS}$$

Can we do better by ML estimate?

- if noise is correlated across time
- if noise variance changes over time

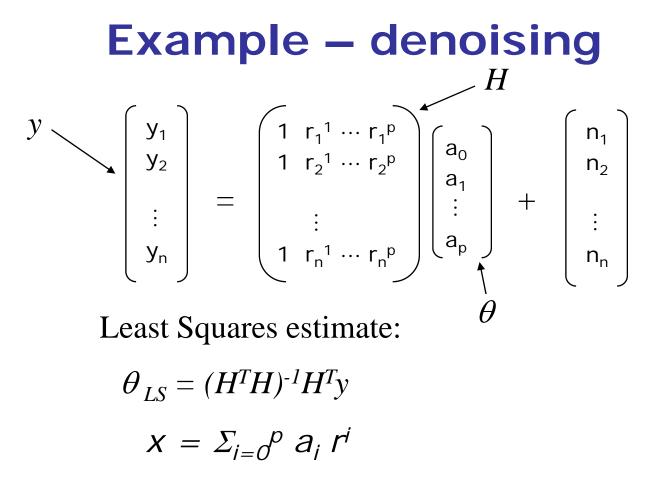
## **Estimation example - Denoising**

 Suppose we have a noisy MR image y, and wish to obtain the noiseless image x, where

y = x + n

- Can we use Estimation theory to find x?
- Try: H = I,  $\theta = x$  in the linear model
- Both LS and ML estimators simply give x = y!
- we need a more powerful model
- Suppose the image x can be approximated by a polynomial, i.e. a mixture of 1<sup>st</sup> p powers of r:

$$\mathbf{x} = \sum_{i=0}^{p} \mathbf{a}_{i} \mathbf{r}^{i}$$



Can we do better by ML estimate? YES Noise in MR can be spatially correlated - ML with covariance matrix C is better

## Multi-variate FLASH

- Acquire 6-10 accelerated FLASH data sets at different flip angles or TR's
- Generate T<sub>1</sub> maps by fitting to:

 $S = \exp\left(-TE/T_2^*\right) \sin \alpha \frac{1 - \exp\left(-TR/T_1\right)}{1 - \cos \alpha \exp\left(-TR/T_1\right)}$ 

- Not enough info in a single voxel
- Noise causes incorrect estimates
- Error in flip angle varies spatially!

# Spatially Coherent T<sub>1</sub>, ρ estimation

- First, stack parameters from all voxels in one big vector x
- Stack all observed flip angle images in y
- Then we can write  $\mathbf{y} = \mathbf{M}(\mathbf{x}) + \mathbf{n}$
- Recall M is the (nonlinear) functional obtained from

$$S = \exp\left(-TE/T_2^*\right) \sin \alpha \frac{1 - \exp\left(-TR/T_1\right)}{1 - \cos \alpha \exp\left(-TR/T_1\right)}$$

Solve for x by non-linear least square fitting, PLUS spatial prior:

$$\mathbf{x}_{est} = \arg \min_{\mathbf{x}} || \mathbf{y} - \mathbf{M}(\mathbf{x}) ||^2 + \mu^2 ||\mathbf{D}\mathbf{x}||^2 \longleftarrow E(\mathbf{x})$$

Makes M(x) close to y

Makes x smooth

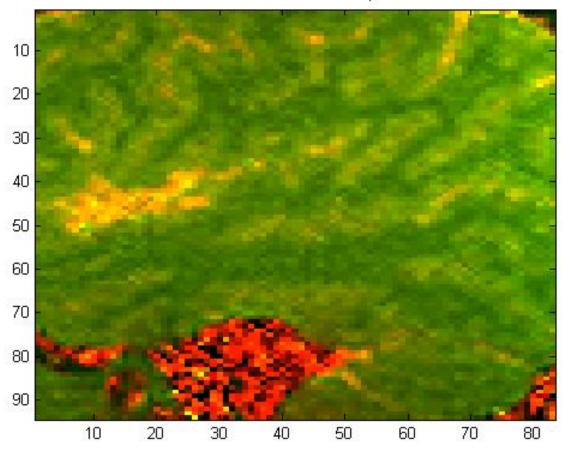
- Minimize via MATLAB's *lsqnonlin* function
- How? Construct  $\delta = [\mathbf{y} \mathbf{M} (\mathbf{x}); \mu D\mathbf{x}]$ . Then  $E(\mathbf{x}) = ||\delta||^2$

### Multi-Flip Results – combined ρ, T<sub>1</sub> in pseudocolour



### Multi-Flip Results – combined ρ, T<sub>1</sub> in pseudocolour

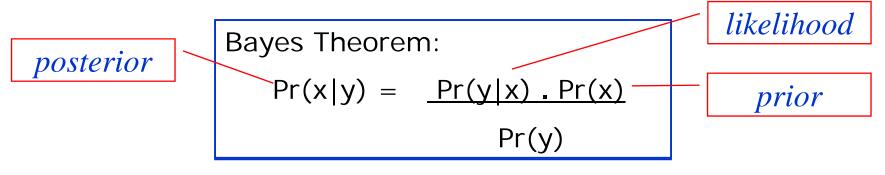
combination T1-rho map



IDEA Lab, Radiology, Cornell

## Maximum a Posteriori Estimate

- This is an example of using an image prior
- Priors are generally expressed in the form of a PDF Pr(x)
- Once the likelihood L(x) and prior are known, we have complete statistical knowledge
- LS/ML are suboptimal in presence of prior
- MAP (aka Bayesian) estimates are optimal



## **Other example of Estimation in MR**

- Image denoising: H = I
- Image deblurring: H = convolution mtx in img-space
- Super-resolution: H = diagonal mtx in k-space
- Metabolite quantification in MRSI



### What Is the Right Imaging Model?

y = Hx + n, n is Gaussian (1)

$$y = Hx + n$$
,  $n, x \text{ are Gaussian}$  (2)  
MAP Sense

#### MAP Sense = Bayesian (MAP) estimate of (2)

### Intro to Bayesian Estimation

- Bayesian methods maximize the posterior probability:  $Pr(x|y) \propto Pr(y|x) \cdot Pr(x)$
- Pr(y|x) (likelihood function) =  $exp(-||y-Hx||^2)$
- *Pr(x)* (prior PDF) = exp(-G(x))
- Gaussian prior:

$$Pr(x) = \exp\{-\frac{1}{2} x^{T} R_{x}^{-1} x\}$$

MAP estimate:

 $x_{est} = arg min ||y-Hx||^2 + G(x)$ 

 MAP estimate for Gaussian everything is known as Wiener estimate

### Regularization = Bayesian Estimation!

- For any regularization scheme, its almost always possible to formulate the corresponding MAP problem
- MAP = superset of regularization

$$Prior model \longrightarrow MAP \longrightarrow Regularization scheme$$

So why deal with regularization??

### Lets talk about Prior Models

- Temporal priors: smooth time-trajectory
- Sparse priors: L0, L1, L2 (=Tikhonov)
- Spatial Priors: most powerful for images
- I recommend robust spatial priors using Markov Fields
- Want priors to be general, not too specific
- Ie, weak rather than strong priors

### How to do regularization

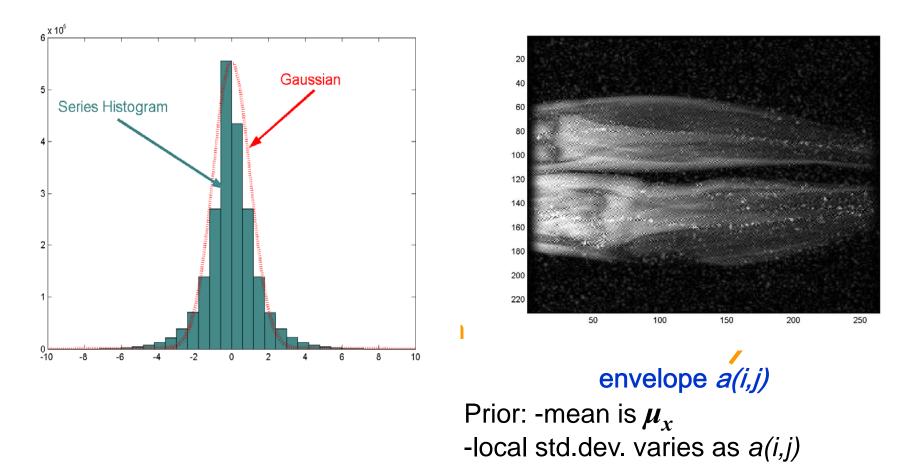
- First model physical property of image,
- then create a prior which captures it,
- then formulate MAP estimator,
- Then find a good algorithm to solve it!

### How NOT to do regularization

- DON'T use regularization scheme without bearing on physical property of image!
- Example: L1 or L0 prior in k-space!
- Specifically: deblurring in k-space (handy b/c convolution becomes multiply)
- BUT: hard to impose smoothness priors in k-space → no meaning

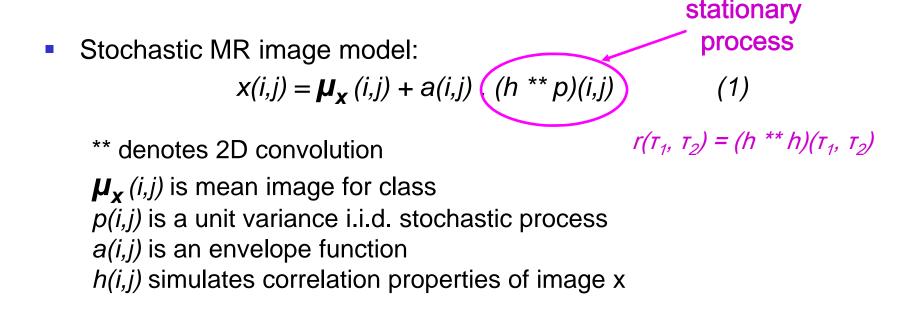
#### **Spatial Priors For Images - Example**

Frames are tightly distributed around mean After subtracting mean, images are close to Gaussian



IDEA Lab, Radiology, Cornell

#### **Spatial Priors for MR images**



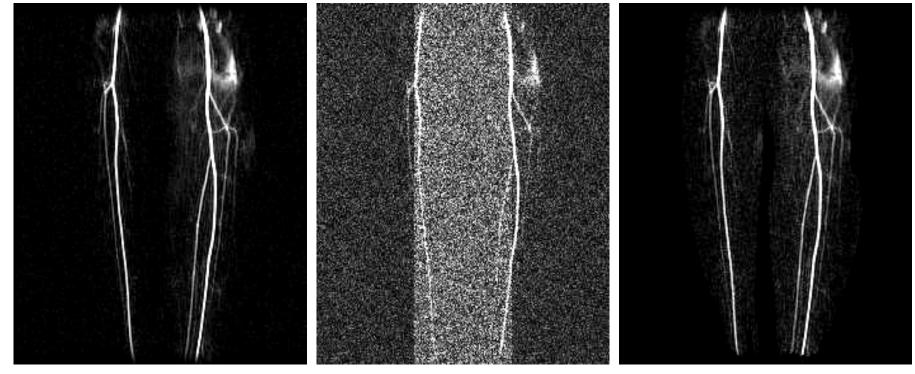
$$x = AC\rho + \boldsymbol{\mu} \tag{2}$$

where A = diag(a), and C is the Toeplitz matrix generated by h

Can model many important stationary and non-stationary cases

#### **MAP-SENSE Preliminary Results**

- Scans acceleraty 5x
- The angiogram was computed by: avg(post-contrast) – avg(pre-contrast)

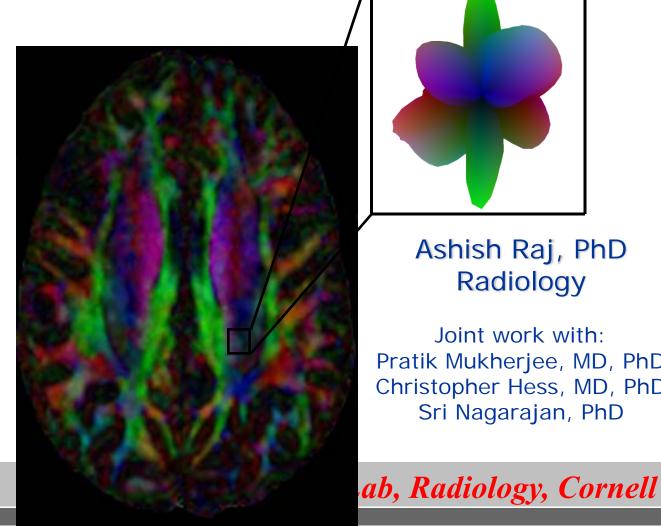


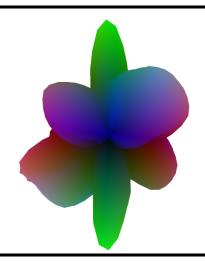
Unaccelerated

5x faster: SENSE

5x faster: MAP-SENSE

#### Spatially Constrained High Angular Resolution Diffusion Imaging





Ashish Raj, PhD Radiology

Joint work with: Pratik Mukherjee, MD, PhD Christopher Hess, MD, PhD Sri Nagarajan, PhD

University of California San Francisco

# **MR Diffusion Imaging**

- Diffusion MRI has revolutionized in vivo imaging of brain
- A new contrast mechanism in addition to T1 or T2
- Measures the directionally varying diffusion properties of water in tissue
- Anisotropy of diffusion is an important marker of extant fiber organization
- Enables non-invasive characterization of white matter integrity
- Enables probing of fiber connectivity in the brain, through tractography

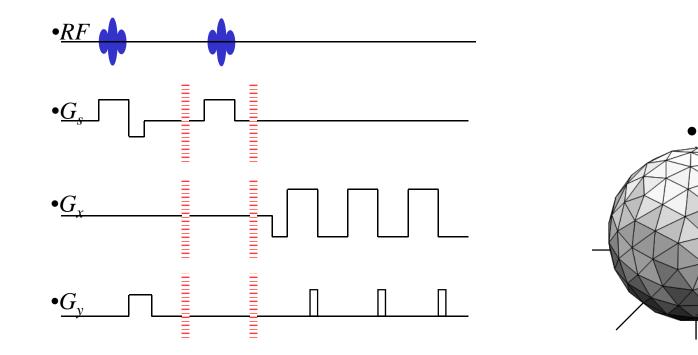
# Diffusion Tensor Imaging (DTI)

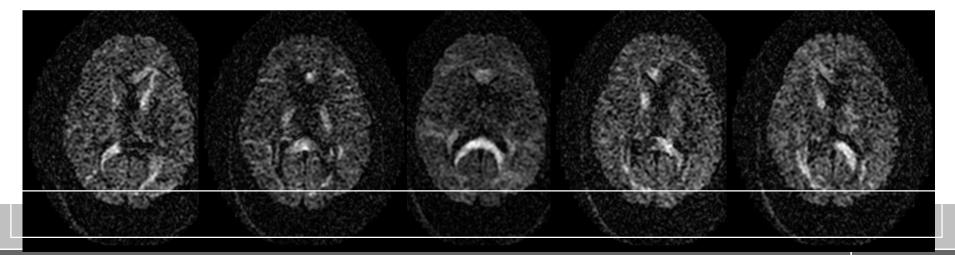
- DTI involves taking 6 directional diffusion imaging measurements
- Then it fits a 3D ellipsoid to these measurements
- Anisotropy of the ellipsoid is correlated with white matter fiber integrity
- Cannot resolve crossing fibers
- Fitting an ellipsoid to crossings gives isotropic spheres
  - Erroneously low FA at crossing fibers
  - Messes up tractography, as well as voxel-wise comparisons
- Need much more than 6 directional measurements to resolve crossing fibers

### • Data Acquisition Strategy

 $\bullet q_y$ 

• $q_x$ 





High Angular Resolution Diffusion Imaging
 Diffusion-encoding Geometries



Gradient directions are determined using an "electrostatic repulsion" model,
for the most uniform sampling of 3D space:
http://www.research.att.com/~njas/electrons/



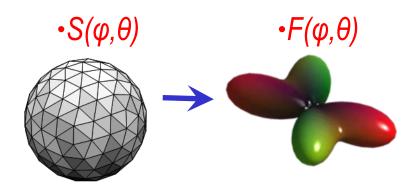
# **Reconstruction Problem**

#### •<u>Goal</u>:

•Construct a spherical function that characterizes the angular structure of diffusion anisotropy in each voxel.

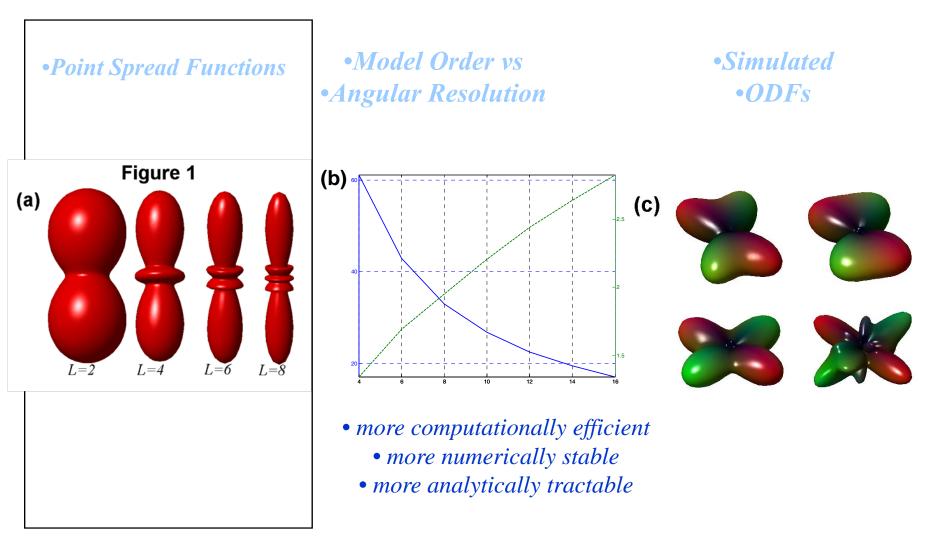
#### •Solutions:

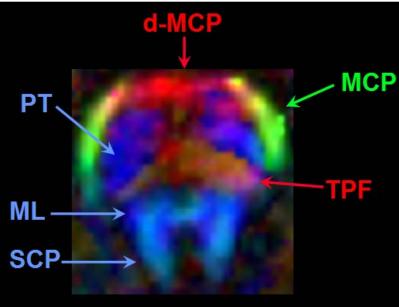
- Multi-tensor fitting
- Generalized DTI
- Persistent angular structure
- Spherical encoding
- Spherical harmonic "ADC profile"
- Circular spectrum mapping
- Spherical deconvolution
- q-ball imaging
- Harmonic q-Ball



Tuch et al, MRM 2002 Özarslan et al, MRM 2003 Jansons et al, Inv. Prob. 2003 Lin et al, ISMRM 2003 Frank, MRM 2002; Alexander DC et al, MRM 2002 Zhan et al, Neuroimage 2004 Tournier et al, Neuroimage 2004 Tuch et al, Neuron 2003; Tuch MRM 2004 Hess et al, ISMRM 2005; MRM 2006

### • High Angular Resolution Diffusion Imaging: Spherical Harmonic Q-ball

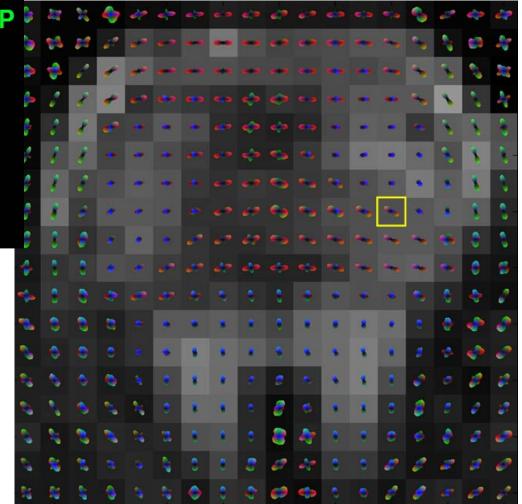




- Middle cerebellar peduncle (MCP)
- Superior cerebellar peduncle (SCP)
- Pyramidal tract (PT)

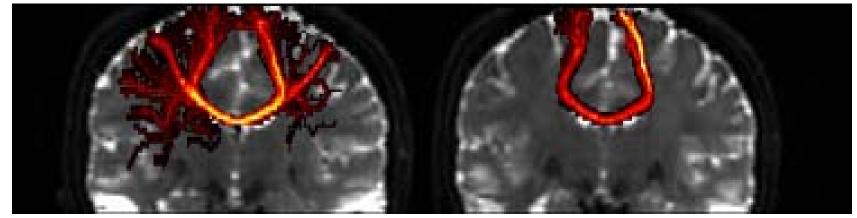
TPF

• Trans pontocerebellar fibers (TPF)



Lab, Radialogy, Cornelloo6; 56:104-117 59

### Clinically Feasible HARDI Tractography



Harmonic q-ball

DTI

60

- Bootstrapping to generate probability distribution function for orientations
- Probabilistic streamline tracking
- 55 direction HARDI protocol, 1x1x2 mm resolution at *b*=3000 s/mm<sup>2</sup>

#### Berman JI, Chung S, Mukherjee P, Hess CP, Han ET, Henry RG. Neuroimage (2007)

# **Problems**

- ODF reconstruction suffers from noise
- Matrix is ill-conditioned
  - i.e. its inverse "magnifies" small noise values into large ones
- There are not enough diffusion directions
- HARDI with high b  $\rightarrow$  very low SNR (<20)

# **Current Solutions**

- Use efficient representation of ODFs
  - Spherical harmonic basis
  - Radial basis
- Limit the order of the basis to use as few basis functions as possible
  - Currently we use spherical harmonics only up to order 4 or 6.
  - Higher order harmonics contain mostly noise
- (but this limits the angular resolution achievable, thus negating the motivation for HARDI)

# Linear System

 Capture the model (whether RBF, SH, ...) into the linear model (matrix) H

y = Hx + n

Then solve for x by inverting H

Usually inversion of H is ill-posed, so add a regularization term

This process is the same for ALL linear estimation problems!

# Regularization

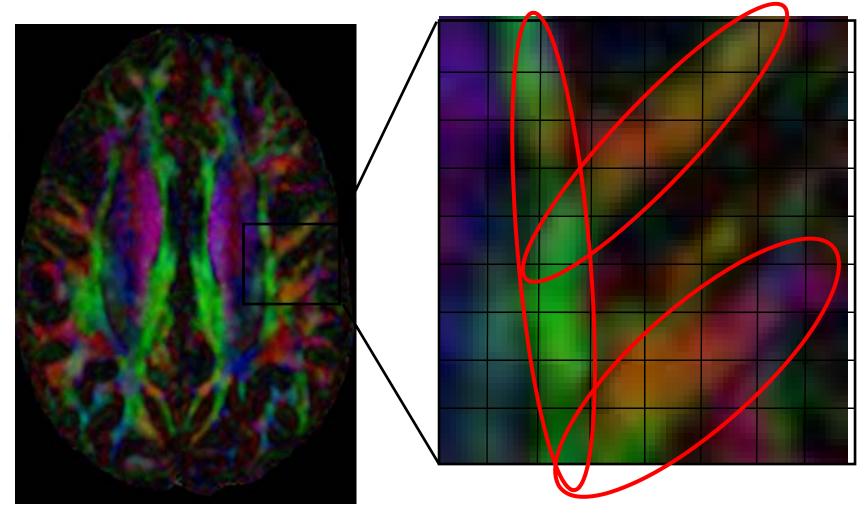
- Regularization of matrix inverse
  - Tikhonov
  - Laplace-Beltrami
- Tikhonov Regularization penalizes all harmonic coefficients
- Laplace-Beltrami penbalizes higher harmonic coefficients more
- Both methods serve to limit the effective angular resolution of reconstructed ODFs

#### A New Approach: add spatial constraints

- Fibers are not arbitrarily arranged in space
- Organized structure follow coherent fiber tracts
- ODFs also have this organized structure
- ODF at one voxel is therefore related to ODF at its neighbours

- 1. How to characterize this neighbourhood relationship?
- 2. How to exploit these spatial constraints to improve ODF reconstruction?

# **Adding Spatial Constraints**

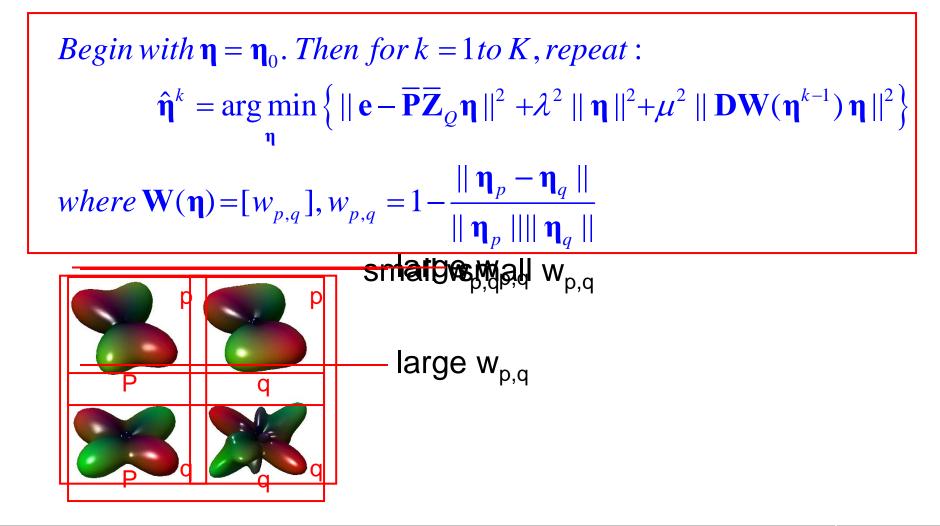


- Neighbours are "like" each other, likely to have similar ODFs
- But need to allow for discontinuous boundaries

# **Iterative Algorithm**

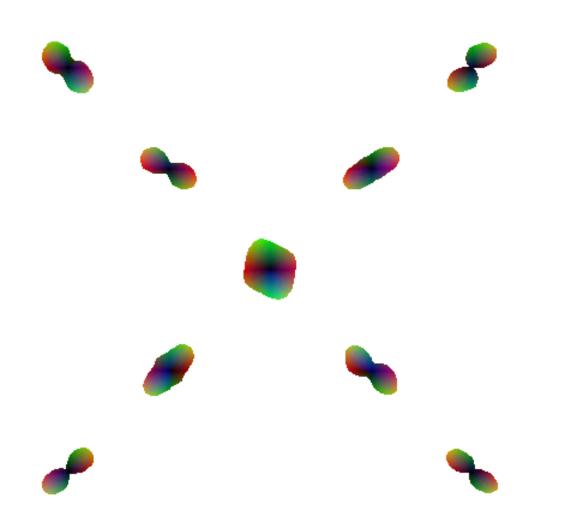
Begin with 
$$\mathbf{\eta} = \mathbf{\eta}_0$$
. Then for  $k = 1$  to  $K$ , repeat :  
 $\hat{\mathbf{\eta}}^k = \arg\min_{\mathbf{\eta}} \left\{ ||\mathbf{e} - \overline{\mathbf{P}}\overline{\mathbf{Z}}_Q \mathbf{\eta}||^2 + \lambda^2 ||\mathbf{\eta}||^2 + \mu^2 ||\mathbf{DW}(\mathbf{\eta}^{k-1})\mathbf{\eta}||^2 \right\}$ 
where  $\mathbf{W}(\mathbf{\eta}) = [w_{p,q}], w_{p,q} = 1 - \frac{||\mathbf{\eta}_p - \mathbf{\eta}_q||}{||\mathbf{\eta}_p|||||\mathbf{\eta}_q||}$ 

# **Iterative Algorithm**

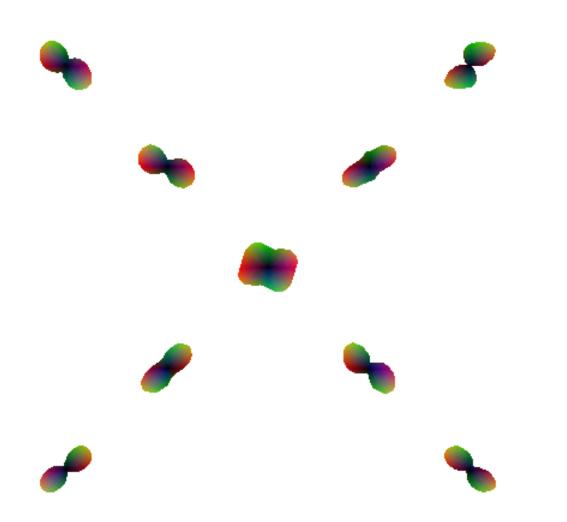


IDEA Lab, Radiology, Cornell

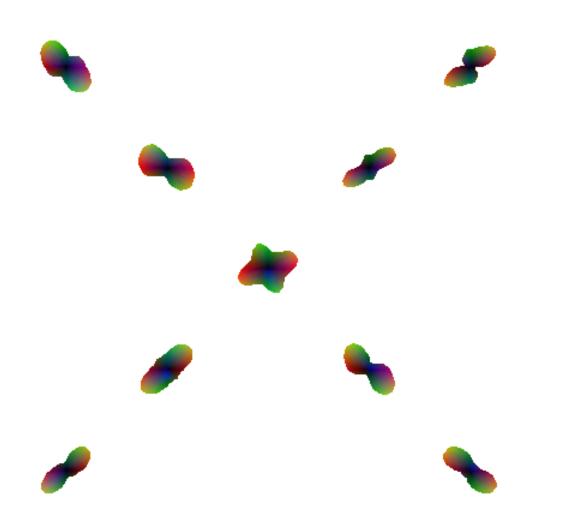
### **Results – simulation**



### **Results – simulation**



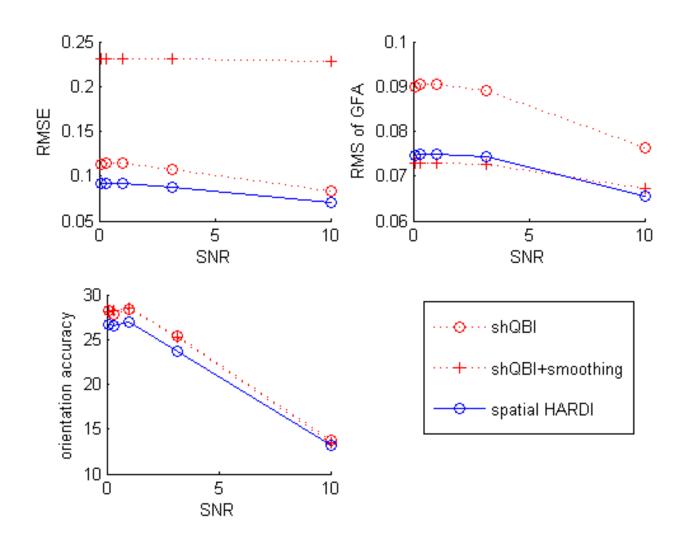
### **Results – simulation**



# Monte Carlo simulations

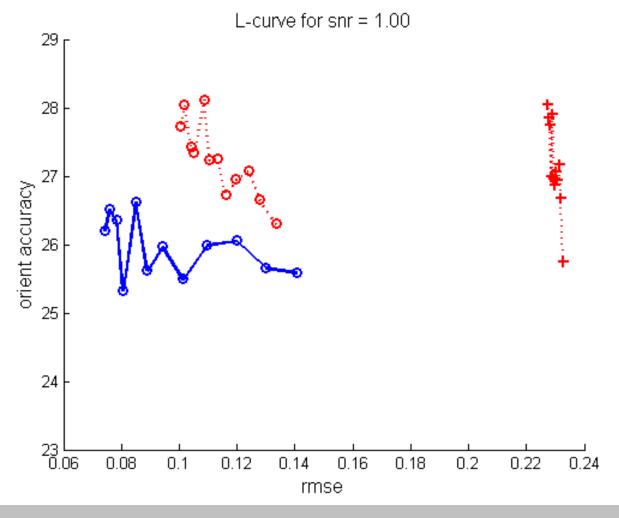
- Repeated multiple times for multiple, random 3D tracts within a 15 x 15 x 15 voxel volume
- Repeated for varying :
  - SNR
  - Algorithm parameters (lambda, mu)
- Evaluation criteria:
  - RMSE
  - Generalized FA
  - Orientation accuracy

### **Monte Carlo simulations**



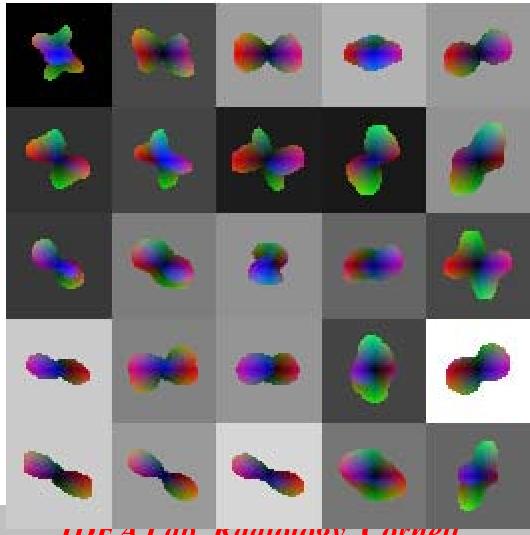
IDEA Lab, Radiology, Cornell

### **Monte Carlo simulations**

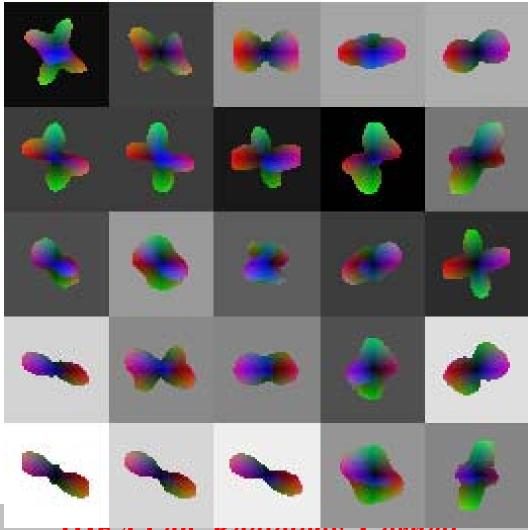


loint IDEA Lab, Radiology, Cornellace

### In vivo results



### In vivo results



#### Part IV : Detection Theory and Examples

Introduction to optimal detection

- Matched filter detectors
- Detection examples from MR

# What is Detection

- Deciding whether, and when, an event occurs
- a.k.a. Decision Theory, Hypothesis testing
- Presence/absence of
  - signal
  - activation (fMRI)
  - foreground/background
  - tissue WM/GM/CSF (segmentation)
- Measures whether statistically significant change has occurred or not

### **Detection**

#### "Spot the Money"



# Hypothesis Testing with Matched Filter

 Let the signal be y(t), model be h(t) Hypothesis testing:

 $H_0: y(t) = n(t)$  (no signal)

 $H_1: y(t) = h(t) + n(t)$  (signal)

 The optimal decision is given by the Likelihood ratio test (Nieman-Pearson Theorem)

Select  $H_1$  if  $L(y) = Pr(y|H_1)/Pr(y|H_0) > \gamma$ 

It can be shown (Kay 01) to be equivalent to
 y(t) \* h(t) > γ'

Matched Filter

IDEA Lab, Radiology, Cornell

# **Matched Filters**

- If the profile of a certain signal is known, it can be detected using the Matched Filter
- If the question is not IF but WHERE...
- Maximum of MF output denotes the most likely location of the object h(t)

# **Matched Filters**

- Example 1: activation in fMRI
  - Need profile model: hemodynamic response function
- Example 2: Detecting malignant tumours in mammograms
  - need profile model: temporal response to contrast agent
- Example 3: Edge detection
- Example 4: detecting contrast arrival in CE-MRA
- In each case need a model to "match" the signal

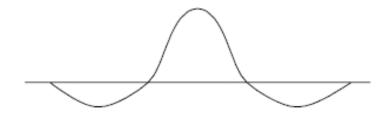
# **Edge Detection**

- Edge information can be used for segmentation
- Detect edges by finding areas of max intensity change
- DoG (Derivative of Gaussian):
   ∇<sup>2</sup> (I(x,y) \* G(x,y,σ))
- $G(x,y,\sigma) = Gaussian$
- $\nabla^2$  = Laplacian operator
- Marr-Hildreth, Canny, Roberts, etc
- Problems: very sensitive to noise, choice of  $\sigma$

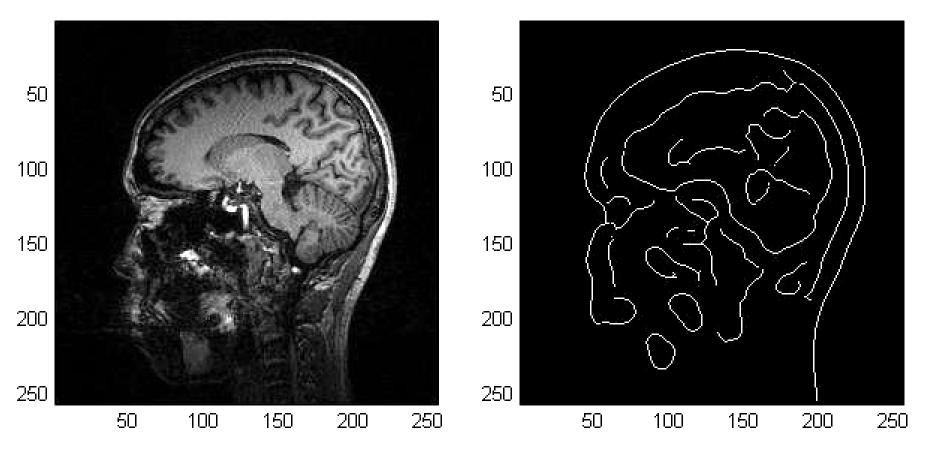
### **Edge Detection**

	1	
1	-4	1
	1	

 $\nabla^2(G_\sigma\star I)$  .



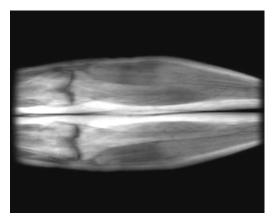
# Edge Detection example using MATLAB

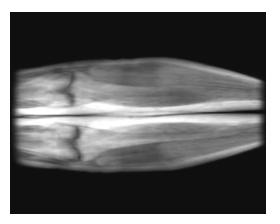


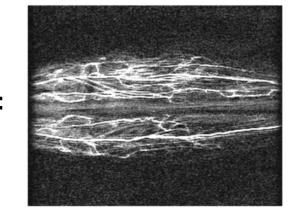
*bw* = *edge*(*I*, *'canny'*, *sigma*);

IDEA Lab, Radiology, Cornell

# Example: Contrast Arrival in CE-MRA





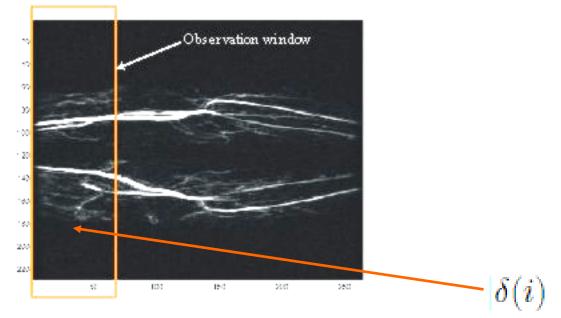


# Mask subtraction in MRA gives vasculature

# Automatic Detection of Contrast Arrival

- MRDSA relies on good estimate of contrast arrival
- Completely unsupervised, reliable automatic method
- >90% accuracy, c.f. earlier reported method (~60% accuracy) matched filter - spatial metric

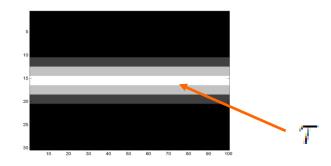
keyhole - frequency metric



Vasculature strongly oriented horizontally



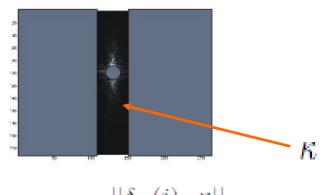
#### **Matched Filter**



$$\lambda_{MF}(i) = \frac{||\delta(i) * *\tau||}{||\tau|| \cdot ||\delta(i)||},$$

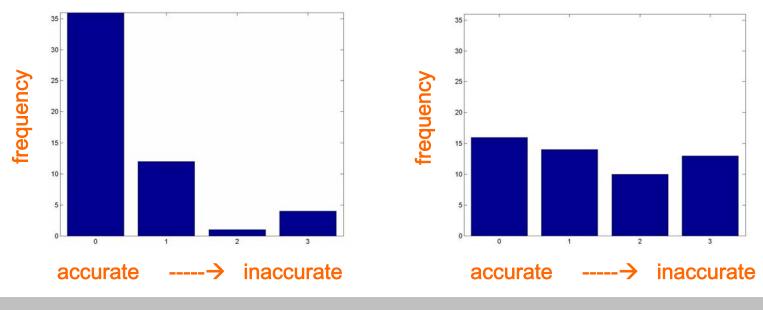
#### Results : Our method





$$\lambda_{KH}(i) = \frac{||\delta_K(i) \cdot \kappa||}{||\delta_K(i)||}.$$

#### Earlier method



IDEA Lab, Radiology, Cornell

# Estimation and Detection of MR Signals

Ashish Raj MR Unit, SFVAMC email: ashished tasheddu ph: 221-4810 x 4800 Visit: http://www.cs.cornell.edu/~rdz/SENSE.htm

This concludes today's lecture. Next week: Classification, Image segmentation, Registration

### References

- Simon Kay. Statistical Signal Processing. Part I: Estimation Theory. Prentice Hall 2002
- Simon Kay. Statistical Signal Processing. Part II: Detection Theory. Prentice Hall 2002
- Haacke et al. Fundamentals of MRI.
- Zhi-Pei Liang and Paul Lauterbur. Principles of MRI A Signal Processing Perspective.

Info on part IV:

- Ashish Raj. Improvements in MRI Using Information Redundancy. PhD thesis, Cornell University, May 2005.
- Website: http://www.cs.cornell.edu/~rdz/SENSE.htm