CS5540: Computational Techniques for Analyzing Clinical Data
Lecture 7:

Statistical Estimation: Least Squares, Maximum Likelihood and Maximum A Posteriori Estimators

Ashish Raj, PhD

Image Data Evaluation and Analytics Laboratory (IDEAL)

Department of Radiology

Weill Cornell Medical College

New York

Outline

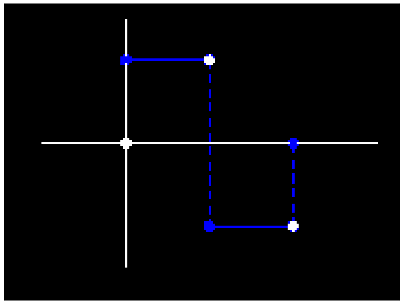
- Part I: Recap of Wavelet Transforms
- Part II : Least Squares Estimation
- Part III: Maximum Likelihood Estimation
- Part IV: Maximum A Posteriori Estimation : Next week

Note: you will not be tested on specific examples shown here, only on general principles

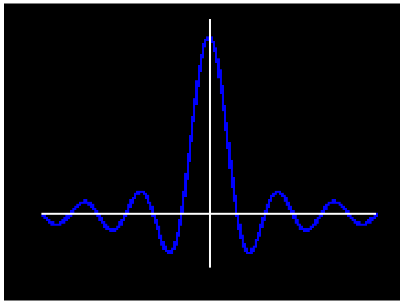
Basis functions in WT

- Basis functions are called "wavelets"
- Important wavelet property:
- All basis functions are scaled, shifted copies of the same mother wavelet
- By clever construction of mother wavelet, these scaled, shifted copies can be made either orthonormal, or at least linearly independent
- Wavelets form a complete basis, and wavelet transforms are designed to be easily invertible
- Online wavelet tutorial:

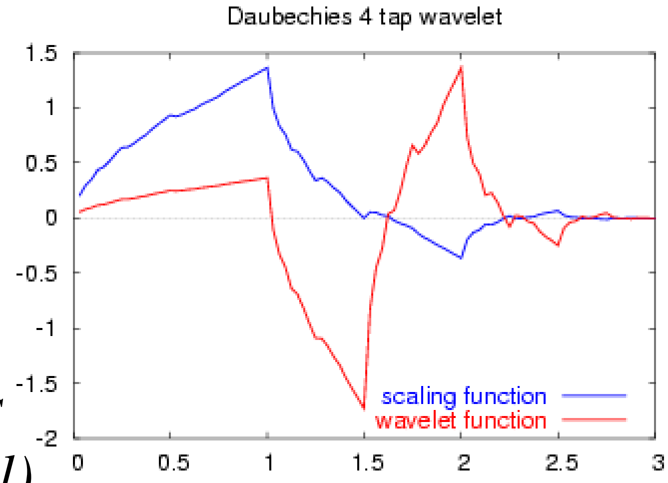
http://cnx.org/content/m10764/latest/



Haar



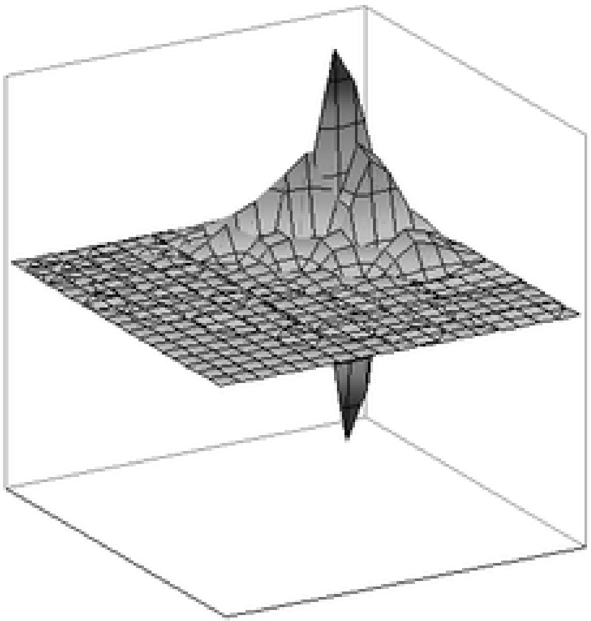
Mexican Hat



Daubechies (orthonormal)

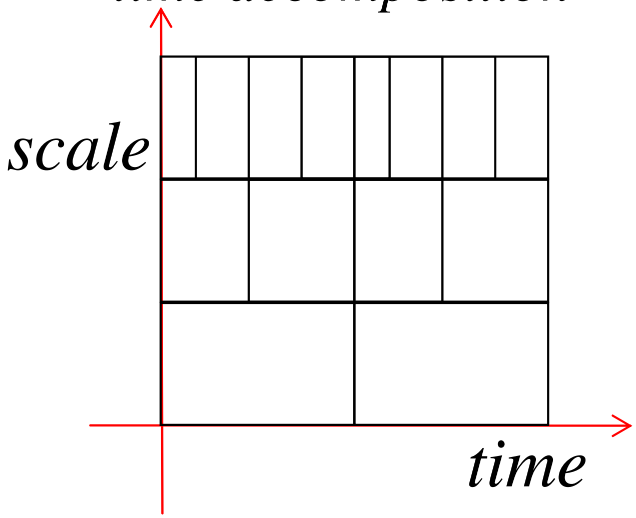
WT in images

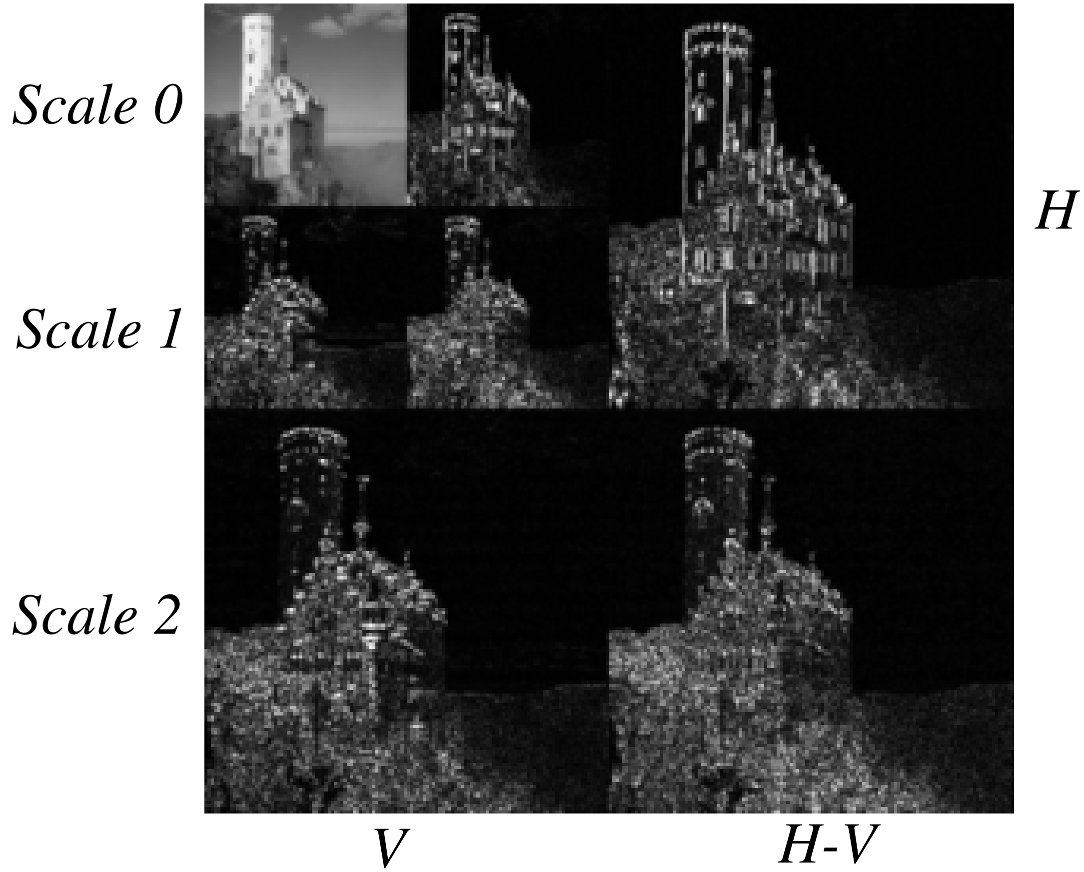
- Images are piecewise smooth or piecewise constant
- Stationarity is even rarer than in 1D signals
- FT even less useful (nnd WT more attractive)
- 2D wavelet transforms are simple extensions of 1D WT, generally performing 1D WT along rows, then columns etc
- Sometimes we use 2D wavelets directly, e.g. orthonormal Daubechies 2D wavelet



WT on images

2D generalization of scale-time decomposition





Successive application of dot product with wavelet of increasing width. Forms a natural pyramid structure. At each scale:

- H = dot product of image rows with wavelet
- V = dot product of image columns with wavelet
- H-V = dot product of image rows then columns with wavelet

Wavelet Applications

- Many, many applications!
- Audio, image and video compression
- New JPEG standard includes wavelet compression
- FBI's fingerprints database saved as wavelet-compressed
- Signal denoising, interpolation, image zooming, texture analysis, time-scale feature extraction
- In our context, WT will be used primarily as a feature extraction tool
- Remember, WT is just a change of basis, in order to extract useful information which might otherwise not be easily seen

WT in MATLAB

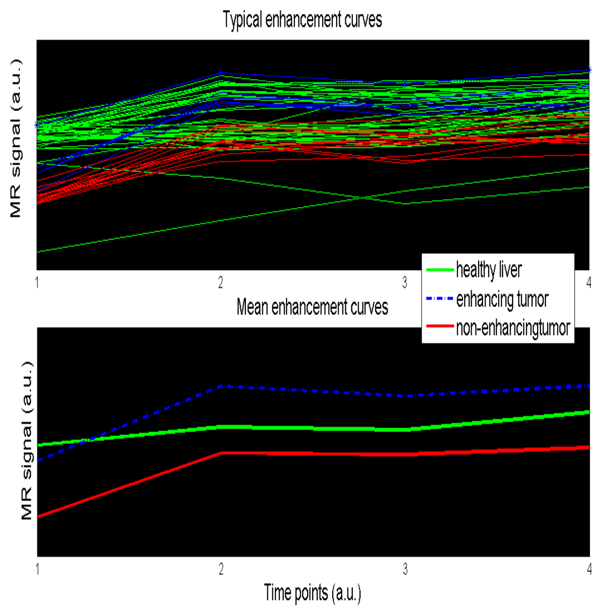
- MATLAB has an extensive wavelet toolbox
- Type help wavelet in MATLAB command window
- Look at their wavelet demo
- Play with Haar, Mexican hat and Daubechies wavelets

Project Ideas

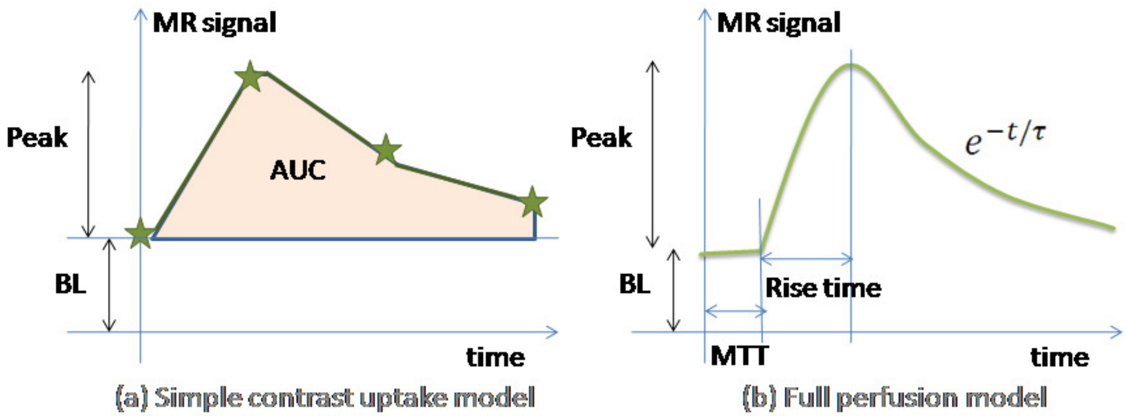
- Idea 1: use WT to extract features from ECG data
  - use these features for classification
- Idea 2: use 2D WT to extract spatio-temporal features from 3D+time MRI data
  - to detect tumors / classify benign vs malignant tumors
- Idea 3: use 2D WT to denoise a given image

Idea 3: Voxel labeling from contrast-enhanced MRI

- Can segment according to time profile of 3D+time contrast enhanced MR data of liver / mammography



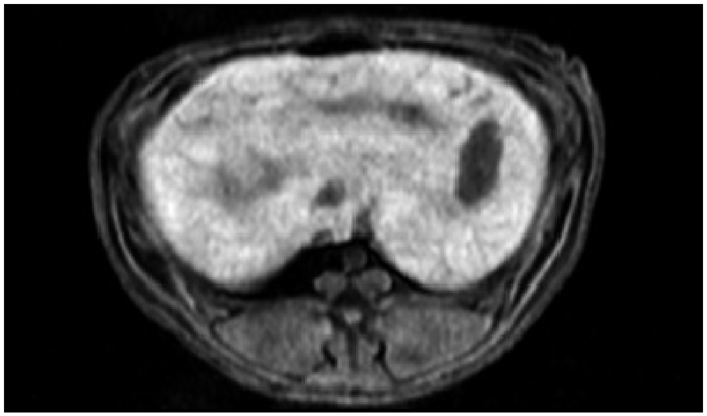
Typical plot of time-resolved MR signal of various tissue classes



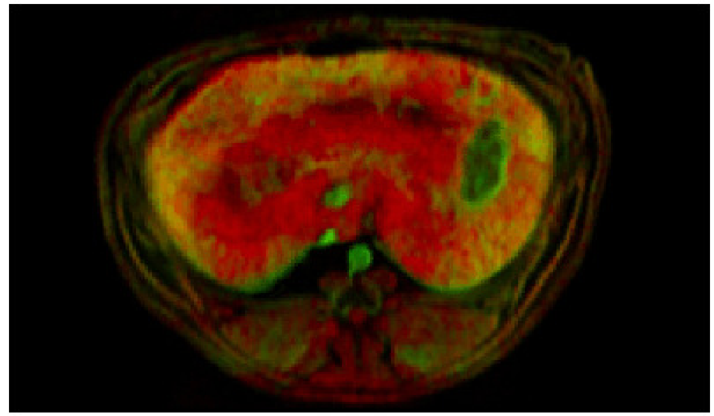
Temporal models used to extract features

Instead of such a simple temporal model, wavelet decomposition could provide spatio-temporal features that you can use for clustering

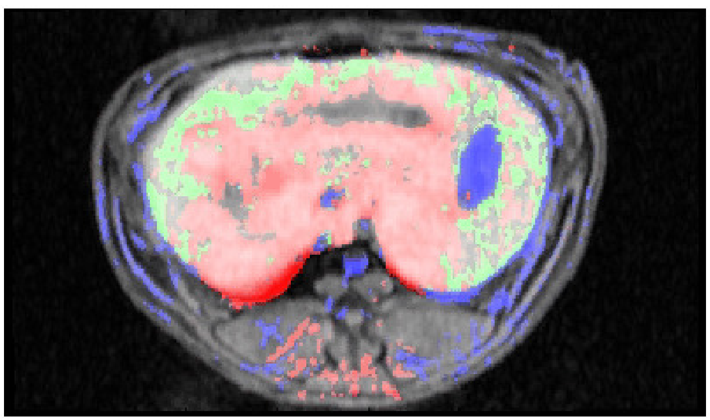
Liver tumour quantification from DCE-MRI



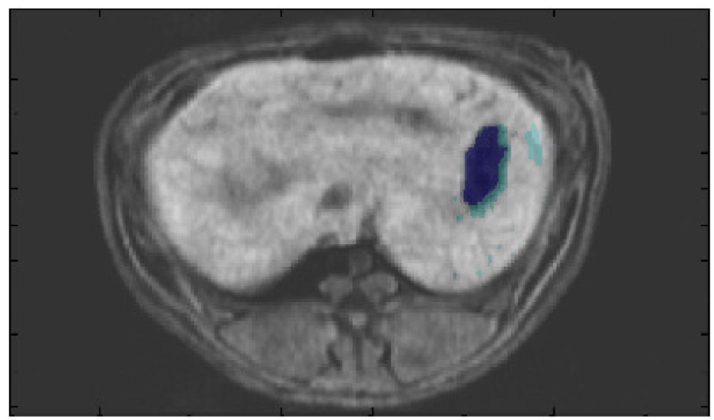
baseline MR image



dynamic parameter map



initial 5-way clustering



final tumor segmentation

Further Reading on Wavelets

- A Linear Algebra view of wavelet transform

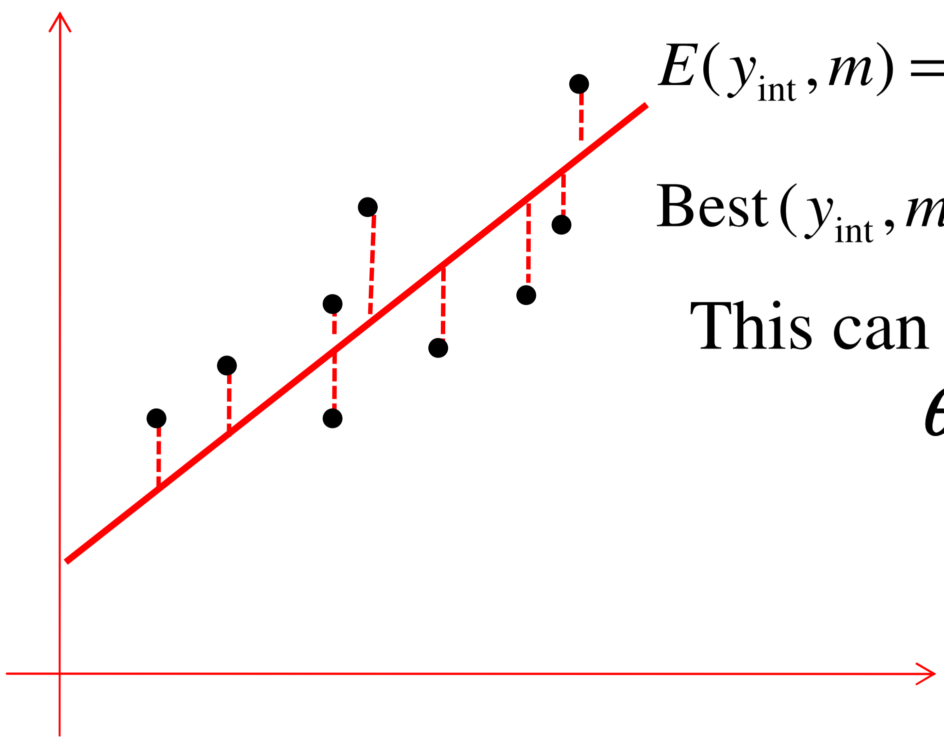
http://www.bearcave.com/misl/misl_tech/wavelets/matrix/index.html

- Wavelet tutorial
- http://users.rowan.edu/~polikar/WAVELETS/WTpart1.html
- http://users.rowan.edu/~polikar/WAVELETS/WTpart2.html
- http://users.rowan.edu/~polikar/WAVELETS/WTpart3.html
- Wavelets application to EKG R wave detection: http://www.ma.utexas.edu/users/davis/reu/ch3/wavelets/wavelets.pdf

Part II : Least Squares Estimation and Examples

A simple Least Squares problem – Line fitting

- Goal: To find the "best-fit" line representing a bunch of points
- Here: y_i are observations at location x_i,
- Intercept and slope of line are the unknown model parameters to be estimated
- Which model parameters best fit the observed points?



$$E(y_{\text{int}}, m) = \sum_i (y_i - h(x_i))^2, \text{ where } h(x) = y_{\text{int}} + m x_i$$

$$\text{Best}(y_{\text{int}}, m) = \arg\min E(y_{\text{int}}, m)$$

This can be written in matrix notation, as

$$\boldsymbol{\theta}_{LS} = arg\ min\ ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}||^2$$

What are $\boldsymbol{H}$ and $\boldsymbol{\theta}$?

Least Squares Estimator

- Given linear process

$$\mathbf{y} = \mathbf{H}\,\theta + \mathbf{n}$$

- Least Squares estimator:

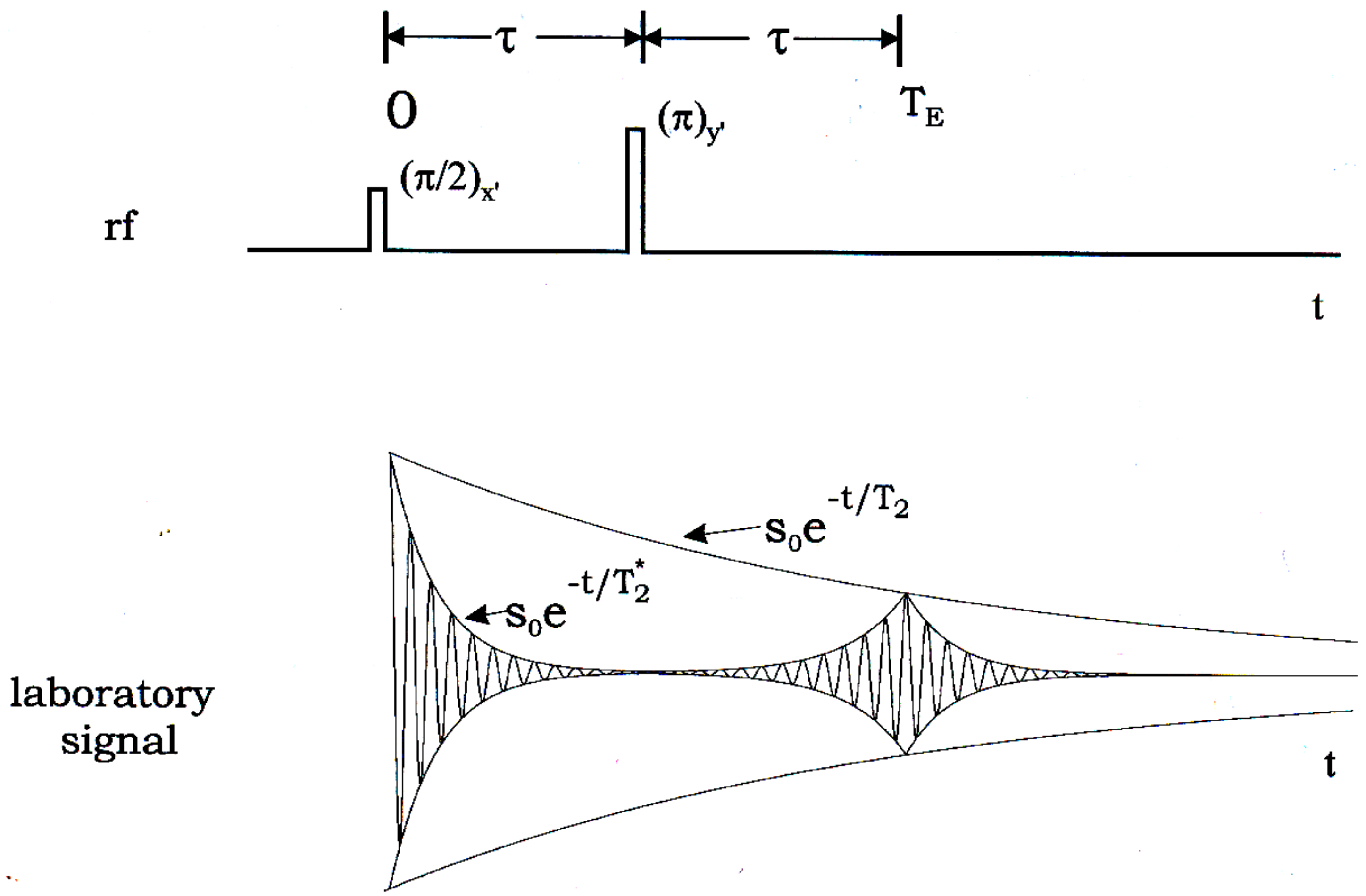
$$\theta_{LS} = \mathrm{argmin}\ ||\mathbf{y} - \mathbf{H}\theta||^2$$

- Natural estimator– want solution to match observation
- Does not use any information about n
- There is a simple solution (a.k.a. pseudo-inverse):

$$\theta_{LS} = (\mathbf{H}^T\mathbf{H})^{-1}\,\mathbf{H}^T\mathbf{y}$$

In MATLAB, type pinv(y)

Example - estimating T_2 decay constant in repeated spin echo MR data



Example – estimating T_2 in repeated spin echo data

$s(t) = s_0 \, e^{-t/T_2}$

- Need only 2 data points to estimate T_2:

  $T_{2est} = [T_{E2} - T_{E1}] \, / \, \ln[s(T_{E1})/s(T_{E2})\,]$
- However, not good due to noise, timing issues
- In practice we have many data samples from various echoes

Example – estimating T_2

$$\underbrace{\begin{pmatrix} \ln(s(t_1)) \\ \ln(s(t_2)) \\ \vdots \\ \ln(s(t_n)) \end{pmatrix}}_{y} = \underbrace{\begin{pmatrix} 1 & -t_1 \\ 1 & -t_2 \\ & \vdots \\ 1 & -t_n \end{pmatrix}}_{H} \underbrace{\begin{bmatrix} a \\ r \end{bmatrix}}_{\theta}$$

Least Squares estimate:

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

$$T_2 = 1/r_{LS}$$

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 LSE to find x?
- Try: $H = I$, $\theta = x$ in the linear model
- LS estimator simply gives $x = y$!

  → we need a more powerful model

- Suppose the image x can be approximated by a polynomial, i.e. a mixture of 1st p powers of r:

  $$x = \Sigma_{i=0}^{p} a_i r^i$$

Example – denoising

$$
\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{y}
=
\underbrace{\begin{pmatrix} 1 & r_1^{1} & \cdots & r_1^{p} \\ 1 & r_2^{1} & \cdots & r_2^{p} \\ & \vdots & & \\ 1 & r_n^{1} & \cdots & r_n^{p} \end{pmatrix}}_{H}
\underbrace{\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix}}_{\theta}
+
\begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_n \end{pmatrix}
$$

Least Squares estimate:

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

$$x = \Sigma_{i=0}^{p} a_i r^i$$

Part III : Maximum Likelihood Estimation and Examples

Estimation Theory

- Consider a linear process

  $$\mathbf{y} = \mathbf{H}\,\theta + \mathbf{n}$$

  $\mathbf{y}$ = observed data

  θ = set of model parameters

  $\mathbf{n}$ = additive noise
- Then Estimation is the problem of finding the statistically optimal θ, given $\mathbf{y}$, $\mathbf{H}$ and knowledge of noise properties
- Medicine is full of estimation problems

Different approaches to estimation

- Minimum variance unbiased estimators
- Least Squares
- Maximum-likelihood
- Maximum entropy
- Maximum a posteriori

has no statistical basis

uses knowledge of noise PDF

uses prior information about θ

Probability vs. Statistics

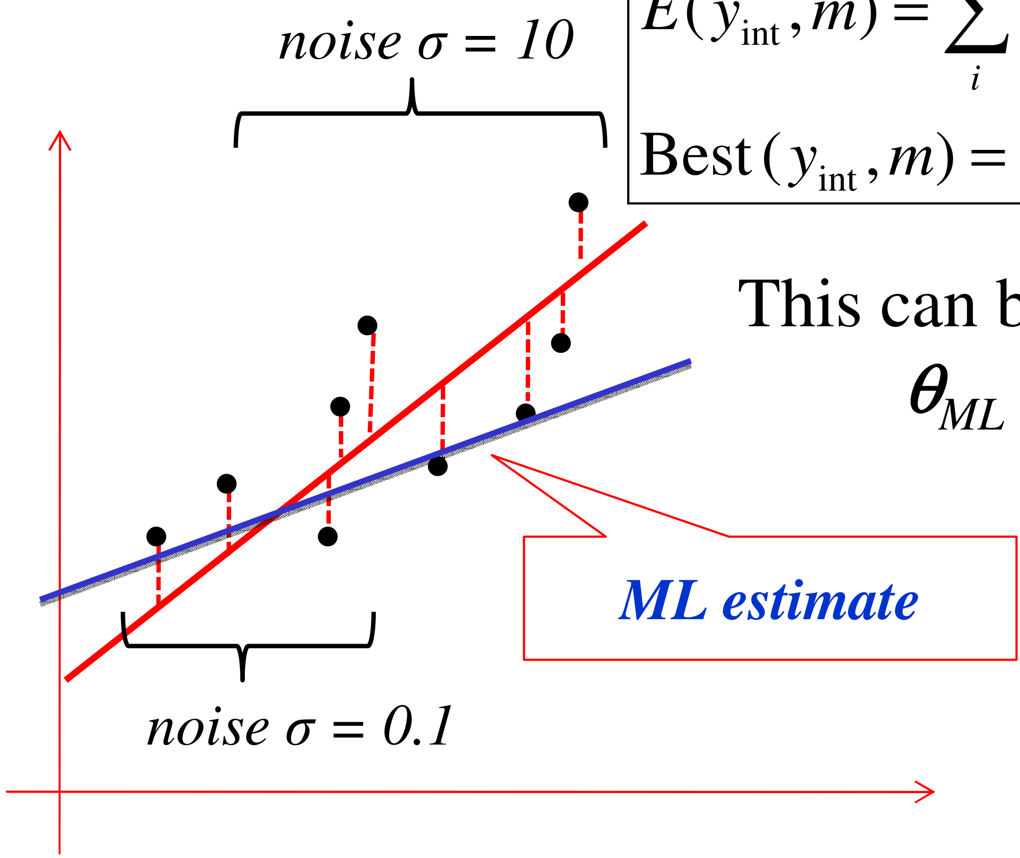
- Probability: Mathematical models of uncertainty predict outcomes
  - This is the heart of probability
  - Models, and their consequences
    - What is the probability of a model generating some particular data as an outcome?
- Statistics: Given an outcome, analyze different models
  - Did this model generate the data?
  - From among different models (or parameters), which one generated the data?

Maximul Likelihood Estimator for Line Fitting Problem

- What if we know something about the noise? i.e. Pr(**n**)…
- If noise not uniform across samples, LS might be incorrect

$$E(y_{\text{int}}, m) = \sum_i \left(\frac{y_i - h(x_i)}{\sigma_i}\right)^2, \text{ where } h(x) = y_{\text{int}} + mx$$

$$\text{Best}(y_{\text{int}}, m) = \arg\min E(y_{\text{int}}, m)$$



This can be written in matrix notation, as

$$\theta_{ML} = arg\ min\ ||\boldsymbol{W}(\mathbf{y} - \boldsymbol{H}\boldsymbol{\theta})||^2$$

What is $\boldsymbol{W}$?

Definition of likelihood

- Likelihood is a probability model of the uncertainty in output given a known input
- The likelihood of a hypothesis is the probability that it would have resulted in the data you saw
  - Think of the data as fixed, and try to chose among the possible PDF's
  - Often, a parameterized family of PDF's
    - ML parameter estimation

Gaussian Noise Models

- In linear model we discussed, likelihood comes from noise statistics
- Simple idea: want to incorporate knowledge of noise statistics
- If uniform white Gaussian noise:

$$\Pr(\mathbf{n}) = \frac{1}{Z}\prod_i \exp\left(-\frac{|n_i|^2}{2\sigma^2}\right) = \frac{1}{Z}\exp\left(-\frac{\sum_i |n_i|^2}{2\sigma^2}\right)$$

- If non-uniform white Gaussian noise:

$$\Pr(\mathbf{n}) = \frac{1}{Z}\exp\left(-\frac{\sum_i |n_i|^2}{2\sigma_i^{\ 2}}\right)$$

Maximum Likelihood Estimator - Theory

- $n = y - H\theta$, $Pr(n) = \exp(-\ ||n||^2/2\sigma^2)$
- Therefore $Pr(y$ for known $\theta) = Pr(n)$
- Simple idea: want to maximize $Pr(y|\theta)$ - called the likelihood function
- Example 1: show that for uniform independent Gaussian noise

$$\theta_{ML} = \arg\min ||y - H\theta||^2$$

- Example 2: For non-uniform Gaussian noise

$$\theta_{ML} = \arg\min ||W(y - H\theta)||^2$$

Maximum Likelihood Estimator

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

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

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

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

- 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

MLE

- Bottomline:
- Use noise properties to write $Pr(y|\theta)$
- Whichever θ maximize above, is the MLE

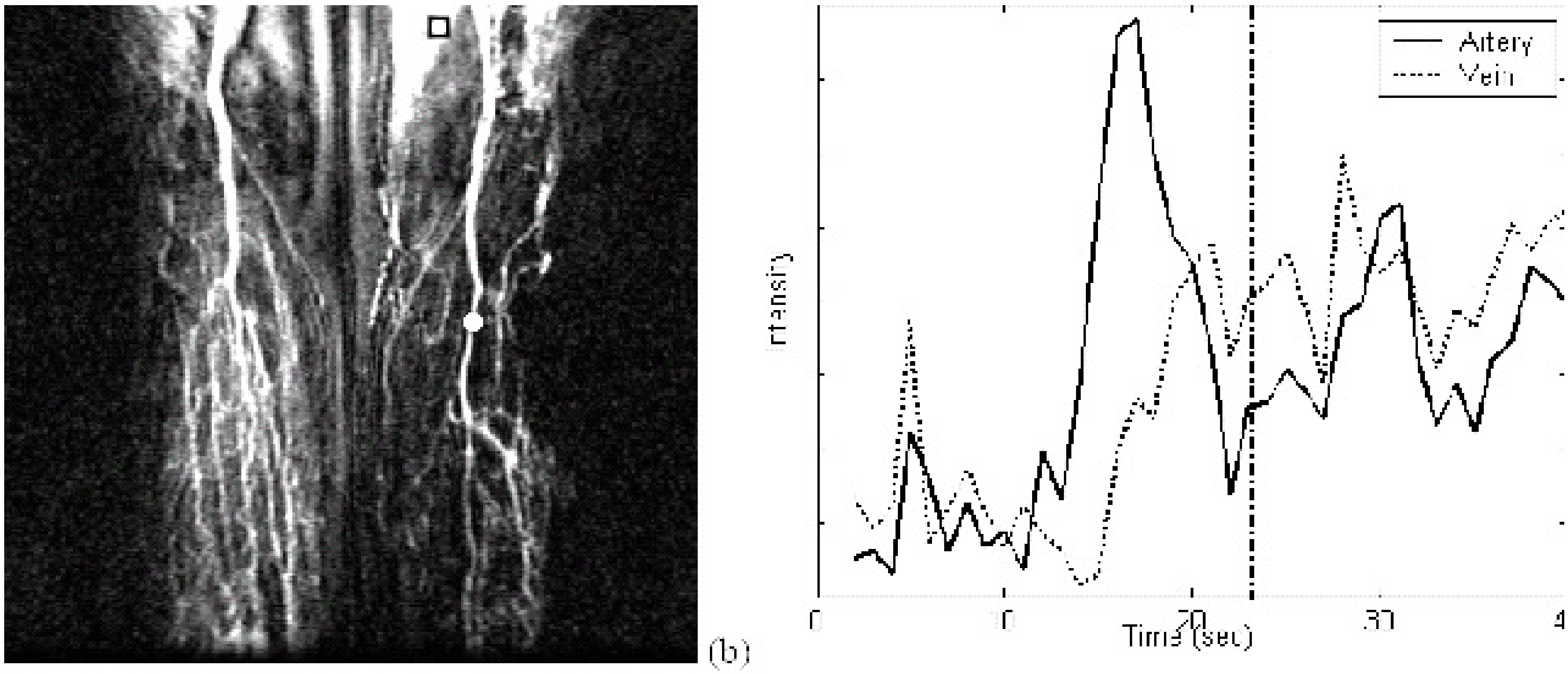
Example – Estimating main frequency of ECG signal

- Model: $y(t_i) = a \sin(f\, t_i) + n_i$
- What is the MLE of a, f ?
- $\Pr(y \mid \theta) = \exp(-\Sigma_i \, (y(t_i) - a \sin(f\, t_i))^2 / 2\,\sigma^2)$

Maximum Likelihood Detection

- ML is quite a general, powerful idea
- Same ideas can be used for classification and detection of features hidden in data
- Example 1: Deciding whether a voxel is artery or vein
- There are 3 hypotheses at each voxel:
  - Voxel is artery, or voxel is vein, or voxel is parenchyma

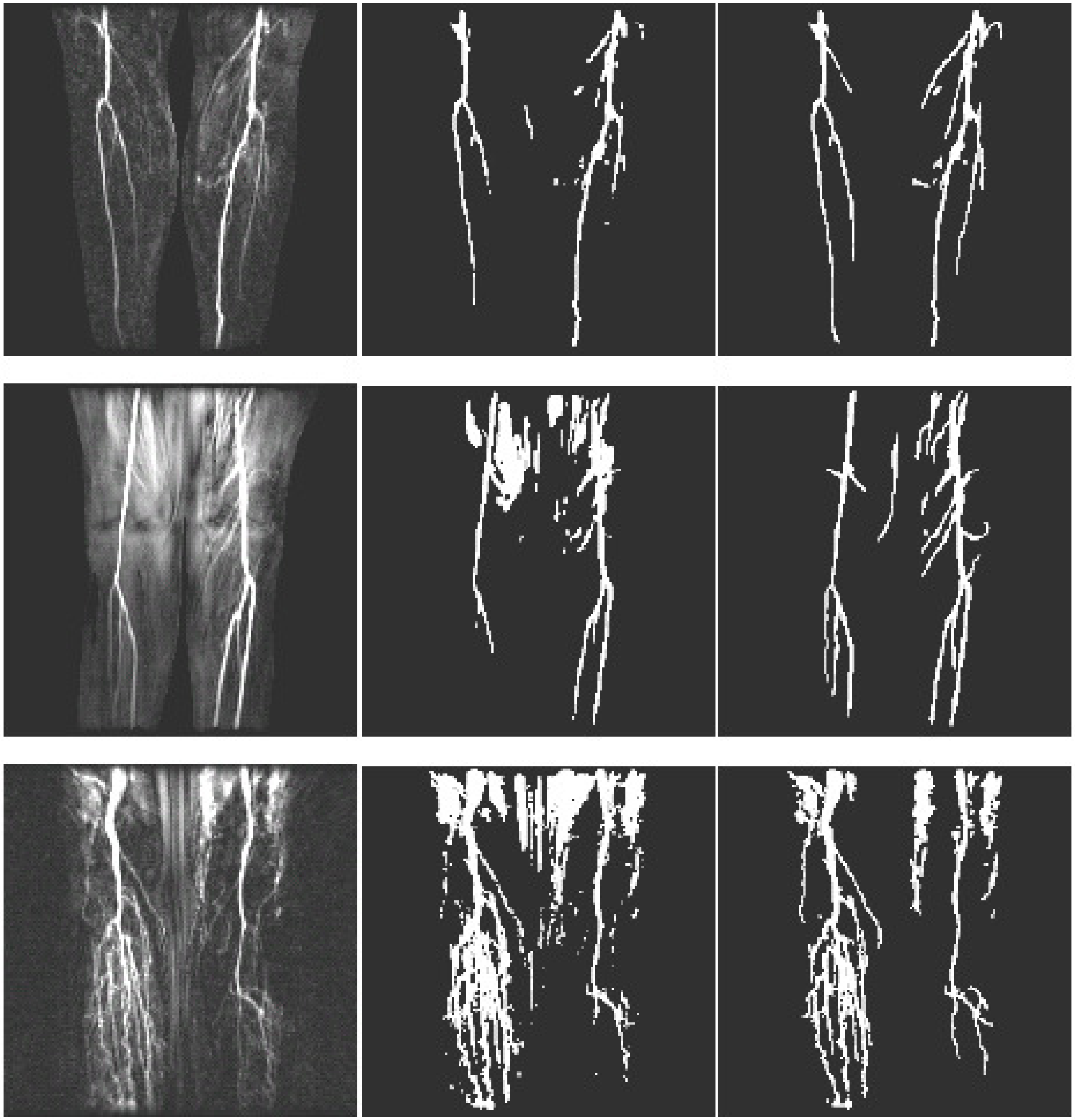
Example: MRA segmentation



(a) (b)

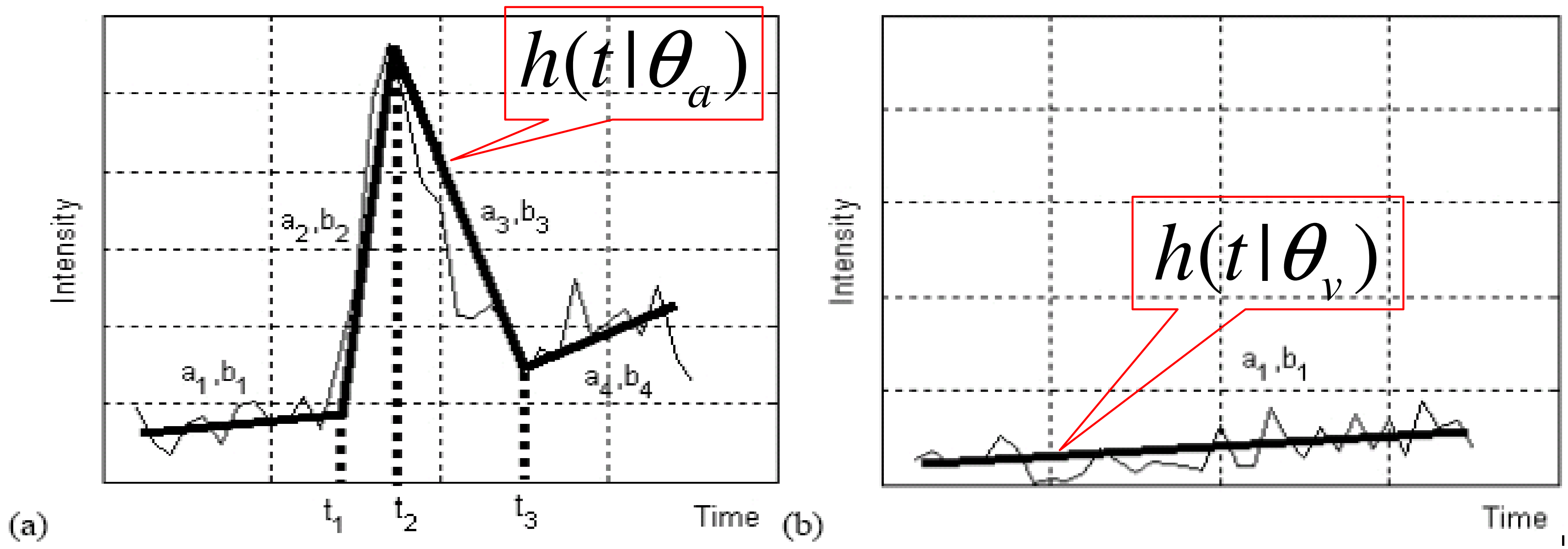
- artery/vein may have similar intensity at given time point
- but different time profiles
- wish to segment according to time profile, not single intensity

Expected Result



Example: MRA segmentation

- First: need a time model of all segments



- Lets use ML principles to see which voxel belongs to which model
- Artery: $y_i = h(t_i | \theta_a) + n_i$
- Vein: $y_i = h(t_i | \theta_v) + n_i$
- Parench: $y_i = h(t_i | \theta_p) + n_i$

Maximum Likelihood Classification

Artery: $y_i = h(t_i \mid \theta_a) + n_i$ $\Rightarrow$ $\Pr(y_i \mid \theta_a) = \exp\left(-\frac{(y_i - h(t_i \mid \theta_a))^2}{2\sigma^2}\right)$

Vein: $y_i = h(t_i \mid \theta_v) + n_i$ $\Rightarrow$ $\Pr(y_i \mid \theta_v) = \exp\left(-\frac{(y_i - h(t_i \mid \theta_v))^2}{2\sigma^2}\right)$

Paren: $y_i = h(t_i \mid \theta_p) + n_i$ $\Rightarrow$ $\Pr(y_i \mid \theta_p) = \exp\left(-\frac{(y_i - h(t_i \mid \theta_p))^2}{2\sigma^2}\right)$

So at each voxel, the best model is one that maximizes:

$$\Pr(y_i \mid \theta) = \prod_i \Pr(y_i \mid \theta) \quad = \exp\left(-\frac{\sum_i (y_i - h(t_i \mid \theta))^2}{2\sigma^2}\right)$$

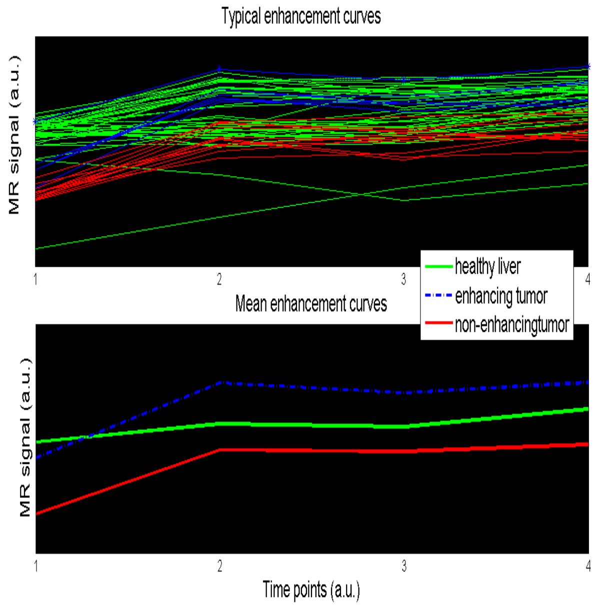
Or equivalently, minimizes:

$$\sum_i (y_i - h(t_i \mid \theta))^2$$

Liver tumour quantification from Dynamic Contrast Enhanced MRI

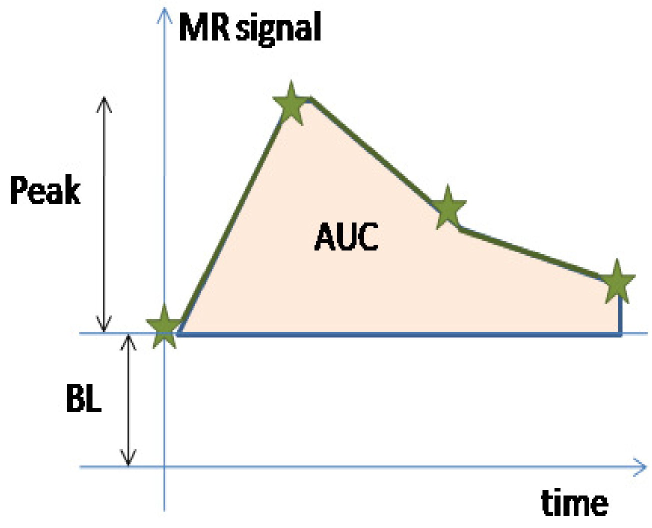
- Data: Tumor model Rabbit DCE-MR data
- Paramegnetic contrast agent , pathology gold standard
- Extract temporal features from DCE-MRI
- Use these features for accurate detection and quantification of tumour

Liver Tumour Temporal models

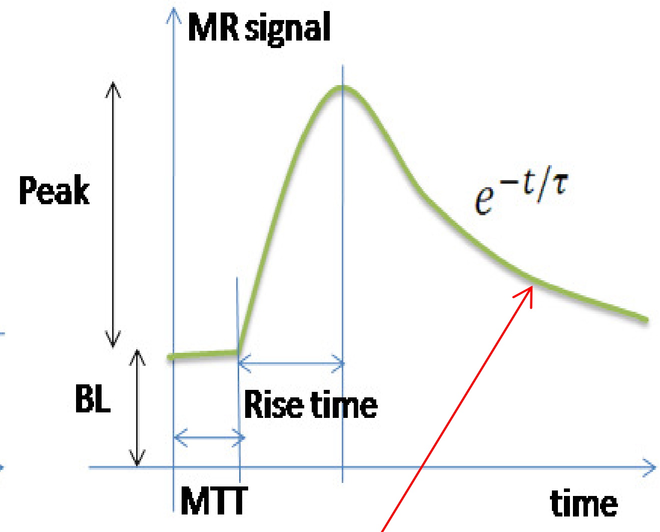


Typical plot of time-resolved MR signal of various tissue classes

Temporal models used to extract features



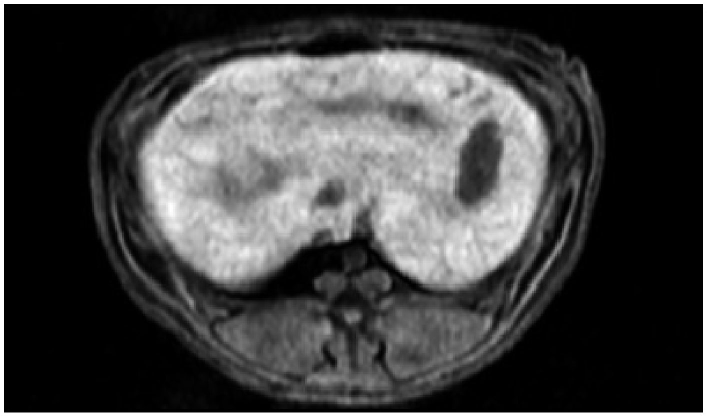
(a) Simple contrast uptake model

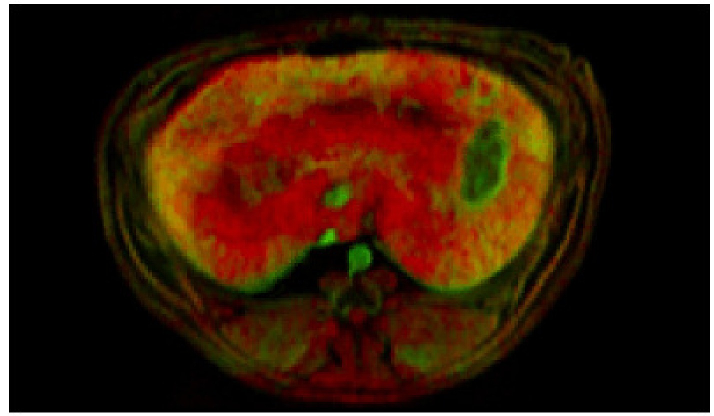


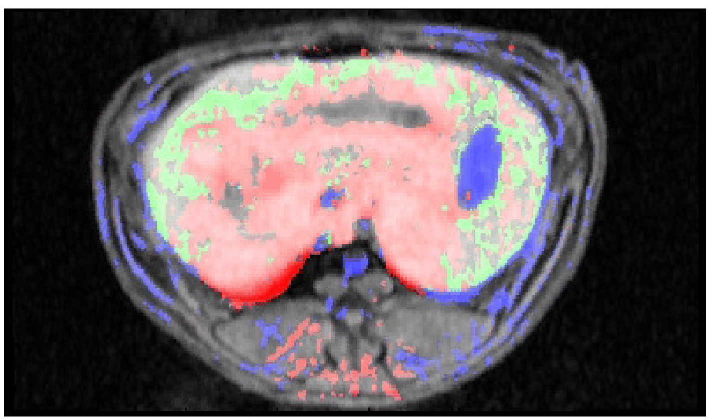
(b) Full perfusion model

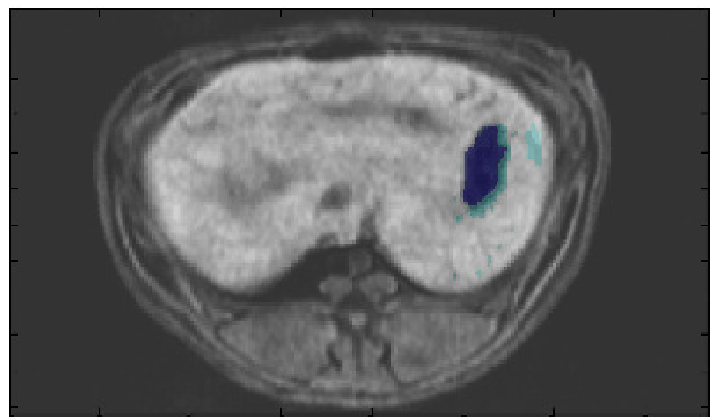
$$\boxed{h(t \mid \theta)}$$

Liver tumour quantification from DCE-MRI


baseline MR image


dynamic parameter map


initial 5-way clustering


final tumor segmentation

ML Tutorials

Slides by Andrew Moore (CMU): available on course webpage

Paper by Jae Myung, “Tutorial on Maximum Likelihood”: available on course webpage

Max Entropy Tutorial

http://www.cs.cmu.edu/~aberger/maxent.html

Part IV : Maximum A Posteriori (Bayesian) Estimation and Examples

Failure modes of ML

- Likelihood isn't the only criterion for selecting a model or parameter
  - Though it's obviously an important one
- ML can be overly sensitive to noise
- Bizarre models may have high likelihood
  - Consider a speedometer reading 55 MPH
  - Likelihood of "true speed = 55": 10%
  - Likelihood of "speedometer stuck": 100%
- ML likes "fairy tales"
  - In practice, exclude such hypotheses
- There must be a principled solution…
  - We need additional information to unwedge ML from bad solutions

If PDF of θ (or x) is also known…

ML

$$\boxed{y = Hx + n, \quad n \text{ is Gaussian} \qquad (1)}$$

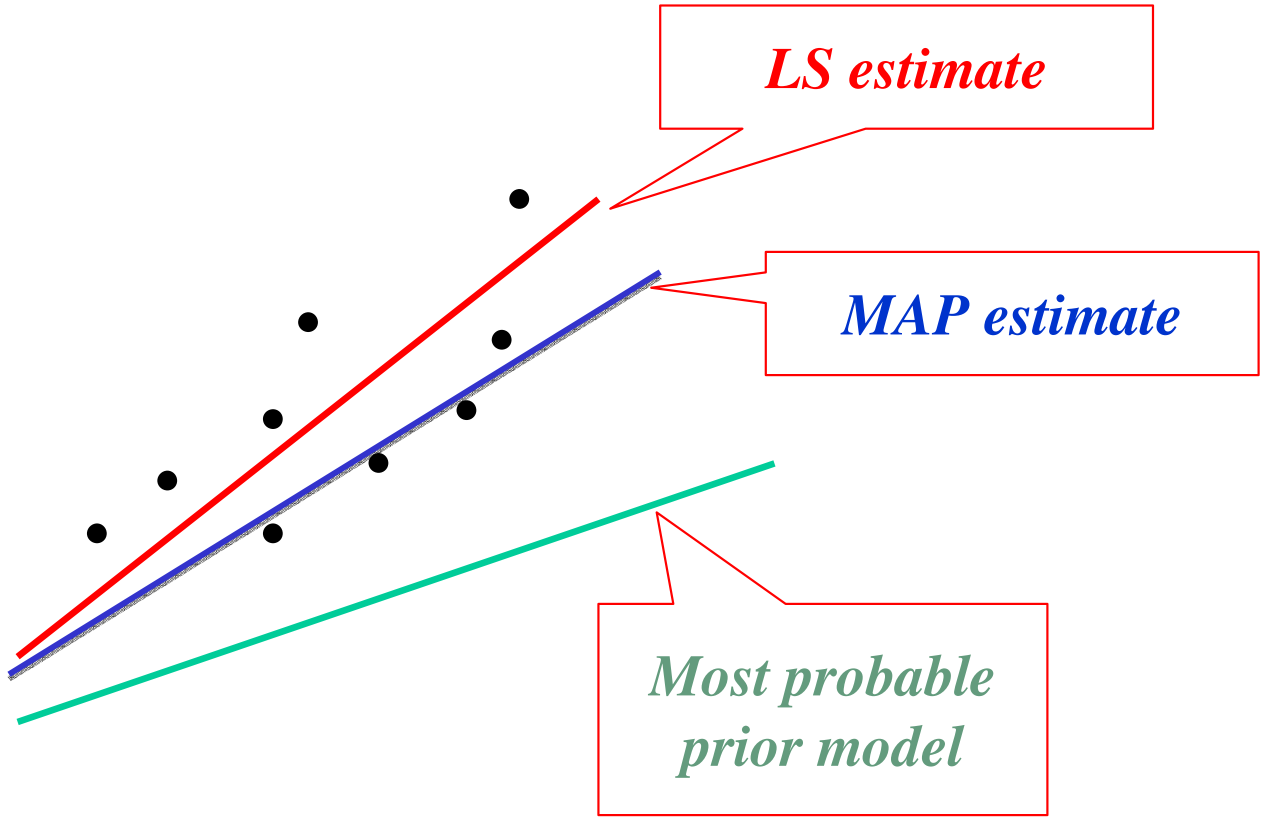
- If we know both Likelihood AND some prior knowledge about the unknown x
- Then can exploit this knowledge
- How? Suppose PDF of x is known

MAP

$$\boxed{y = Hx + n, \quad n, x \text{ are Gaussian} \qquad (2)}$$

MAP for Line fitting problem

- If model estimated by ML and Prior info do not agree…
- MAP is a compromise between the two



Maximum a Posteriori Estimate

- Prior knowledge about random variables is generally expressed in the form of a PDF Pr(x)
- Once the likelihood L(x) and prior are known, we have complete statistical knowledge
- MAP (aka Bayesian) estimates are optimal

posterior

Bayes Theorem:

$$\Pr(x|y) = \frac{\Pr(y|x) \cdot \Pr(x)}{\Pr(y)}$$

likelihood

prior

Maximum a Posteriori (Bayesian) Estimate

- Consider the class of linear systems $y = Hx + n$
- Bayesian methods maximize the posterior probability:

$$Pr(x|y) \propto Pr(y|x)\ .\ Pr(x)$$

- $Pr(y|x)$ (likelihood function) = exp(- $||y\text{-}Hx||^2$)
- $Pr(x)$ (prior PDF) = exp(-G(x))
- Non-Bayesian: maximize only likelihood

$$x_{est} = arg\ min\ ||y\text{-}Hx||^2$$

- Bayesian:

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

where $G(x)$ is obtained from the prior distribution of x

Example Bayesian Estimation

- Example: Gaussian prior centered at zero:

$$\Pr(x) = \exp\{-\tfrac{1}{2}\ x^T R_x^{-1} x\}$$

- Bayesian methods maximize the posterior probability:

$$Pr(x|y)\ \square\ Pr(y|x)\ .\ Pr(x)$$

- $Pr(y|x)$ (likelihood function) $= \exp(-\ ||y-Hx||^2)$
- ML estimate: maximize only likelihood

$$x_{est} = arg\ min\ ||y-Hx||^2$$

- MAP estimate:

$$x_{est} = arg\ min\ ||y-Hx||^2 + \lambda\ x^T R_x^{-1} x$$

Makes Hx close to y

Tries to make x a sample from a Gaussian centered at 0

MAP Example: Multi-variate FLASH

- Acquire 6-10 accelerated FLASH data sets at different flip angles or TR's
- Generate T_1 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_1, 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 **y** = **M** (**x**) + **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:

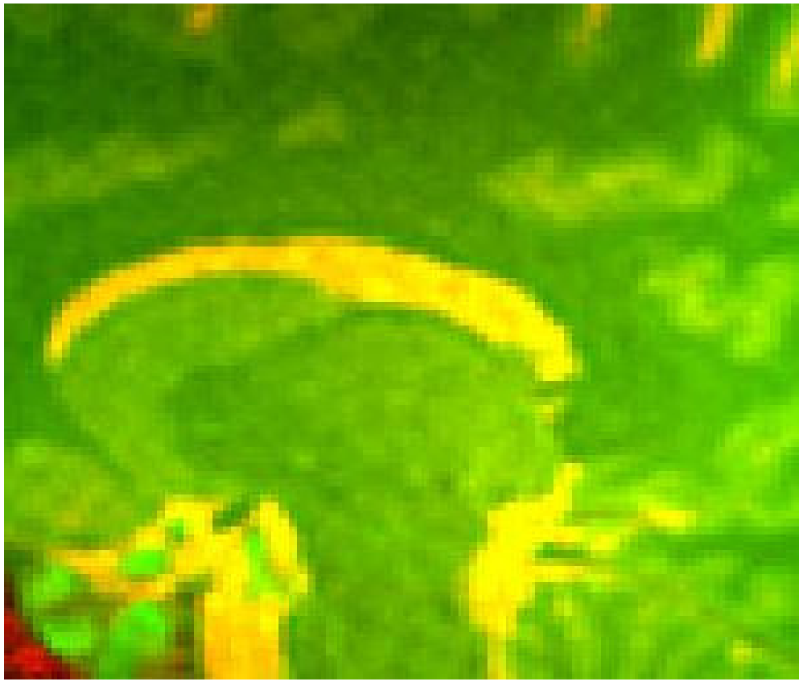
$$\mathbf{x}_{est} = \arg\min_{\mathbf{x}} \boxed{\|\mathbf{y} - \mathbf{M}(\mathbf{x})\|^2 + \mu^2\|D\mathbf{x}\|^2} \Leftarrow E(\mathbf{x})$$

Makes M(x) close to y

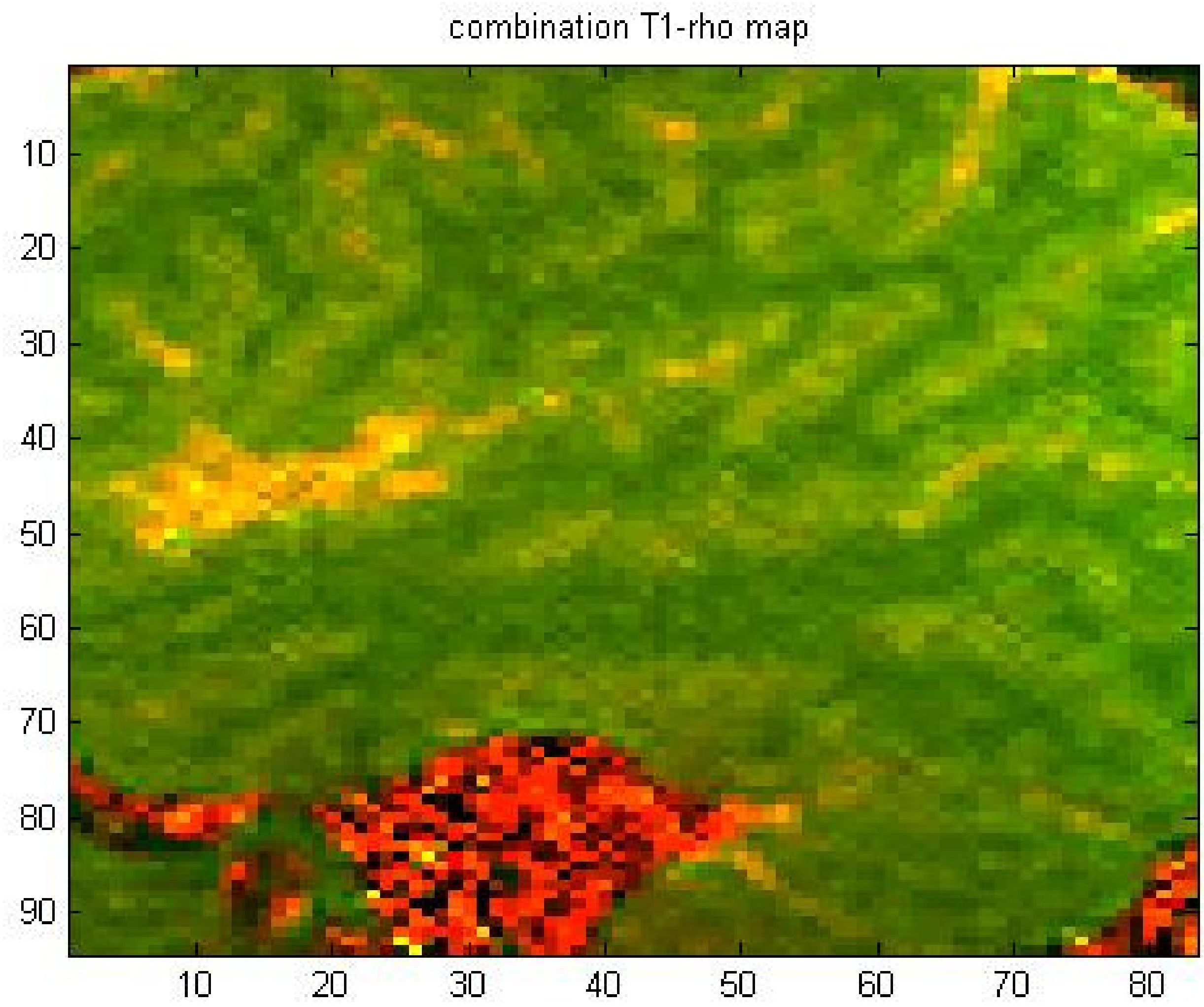
Makes x smooth

- Minimize via MATLAB's *lsqnonlin* function

Multi-Flip Results – combined ρ, T_1 in pseudocolour



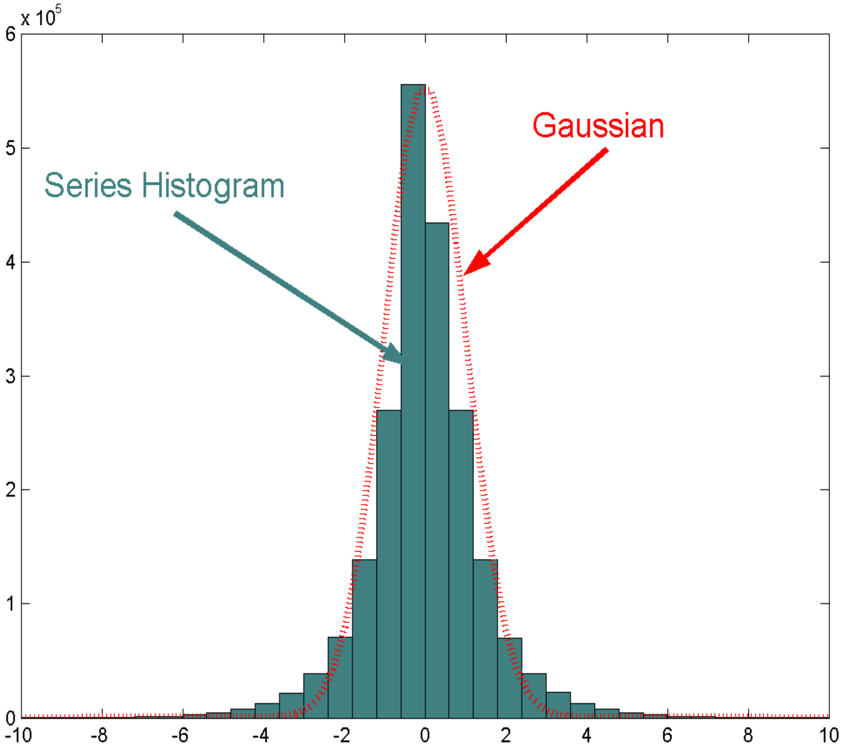
Multi-Flip Results – combined ρ, T_1 in pseudocolour

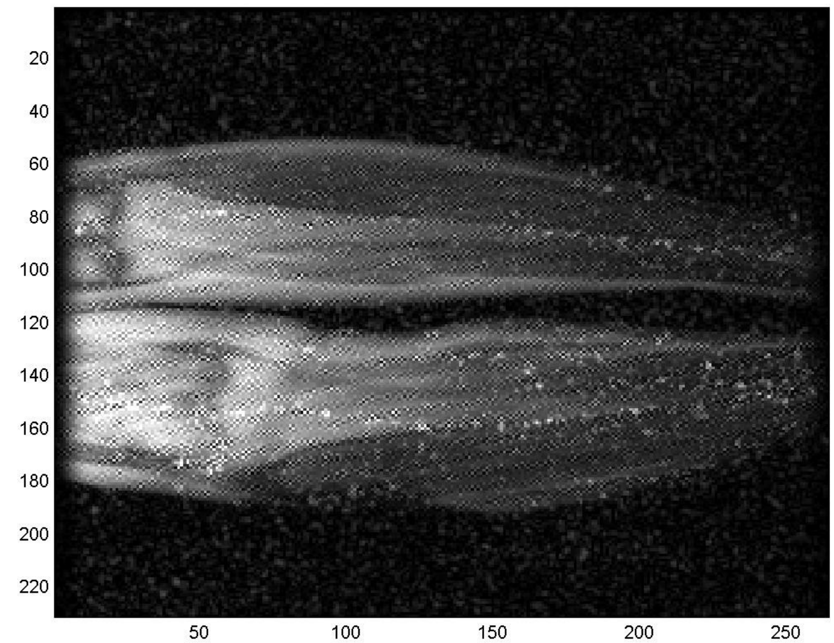


MAP example 2: Spatial Priors For Dynamic Imaging

Frames are tightly distributed around mean

After subtracting mean, images are close to Gaussian





envelope *a(i,j)*

Prior: -mean is $\boldsymbol{\mu_x}$

-local std.dev. varies as *a(i,j)*

Spatial Priors for MR images

stationary process

- Stochastic MR image model:

$$x(i,j) = \boldsymbol{\mu}_{\boldsymbol{x}}(i,j) + a(i,j)\,(h ** p)(i,j) \qquad (1)$$

$r(\tau_1, \tau_2) = (h ** h)(\tau_1, \tau_2)$

** denotes 2D convolution

$\boldsymbol{\mu}_{\boldsymbol{x}}(i,j)$ is mean image for class
p(i,j) is a unit variance i.i.d. stochastic process
a(i,j) is an envelope function
h(i,j) simulates correlation properties of image x

$$x = ACp + \boldsymbol{\mu} \qquad (2)$$

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

- Can model many important stationary and non-stationary cases

MAP estimate for Imaging Model (3)

- The Wiener estimate

$$x_{MAP} - \boldsymbol{\mu}_x = HR_x\,(HR_xH^H + R_n)^{-1}\,(y - \boldsymbol{\mu}_y) \qquad (3)$$

R_x, R_n = covariance matrices of x and n

Stationarity ➔ R_x has Toeplitz structure ➔ fast processing

$$x_{MAP} - \boldsymbol{\mu}_x = HACC^HA^H\;(HACC^HA^HH^H + \sigma_n^2 I)^{-1}\,(y - \boldsymbol{\mu}_y) \quad (4)$$

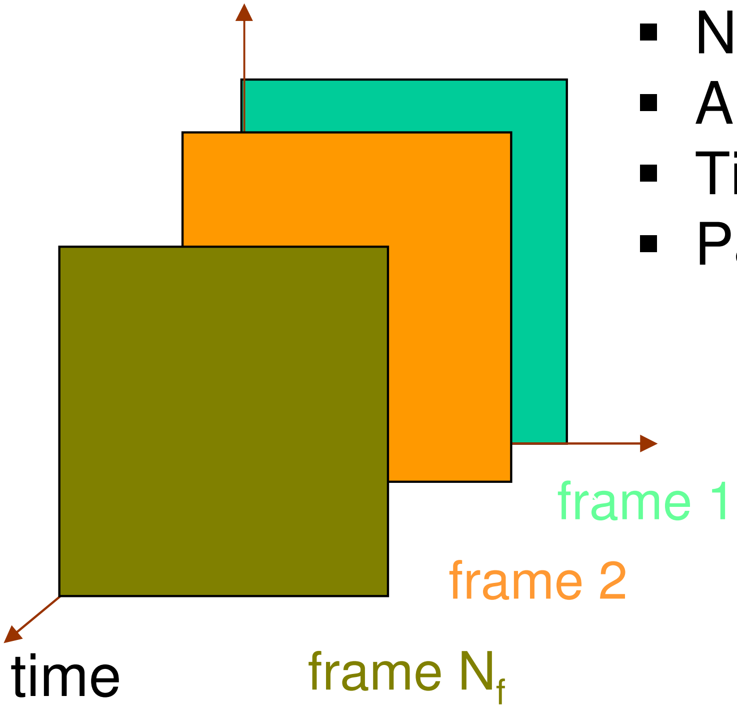
- Direct inversion prohibitive; so use CG iterative method
- (4) better than (3) since A and C are $O(N \log N)$ operations, enabling much faster processing

How to obtain estimates of A, C ?

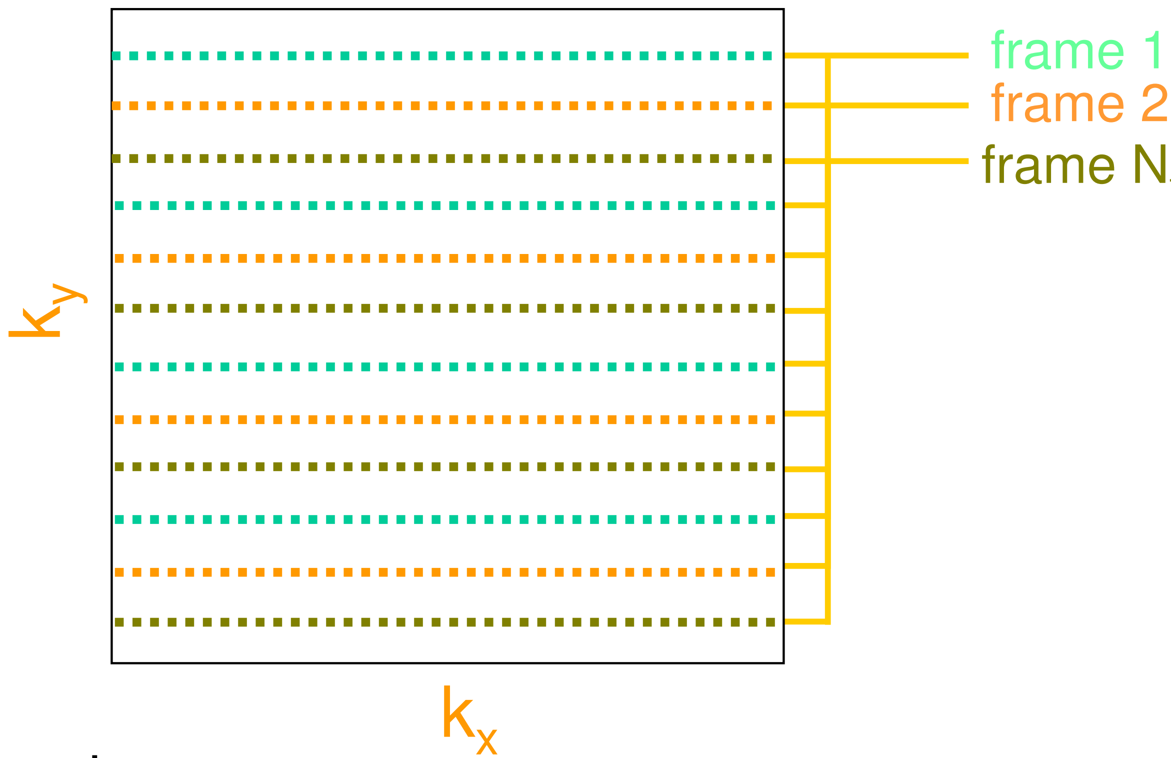
- Need a training set of full-resolution images x_k, k = 1,…,K
- Parallel imaging doesnt provide un-aliased full-res images

- Approaches:
  1. Use previous full-res scans
     - time consuming, need frequent updating
  2. Use SENSE-reconstructed images for training set
     - very bad noise amplification issues for high speedu ps
  3. Directly estimate parameters from available parallel data
     - Aliasing may cause inaccuracies

MAP for Dynamic Parallel Imaging



- N_f images available for parameter estimation!
- All images well-registered
- Tightly distributed around pixel mean
- Parameters can be estimated from aliased data

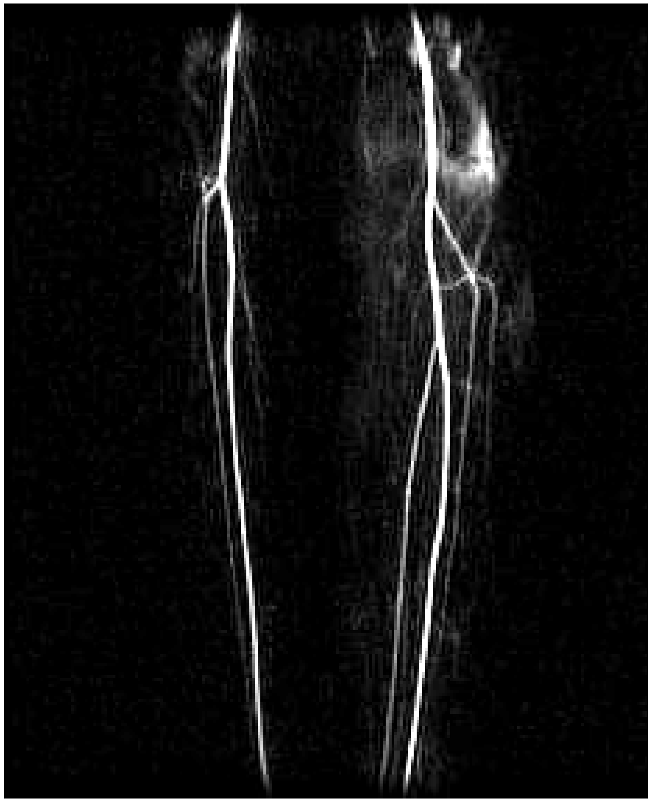


- N_f /R full-res images

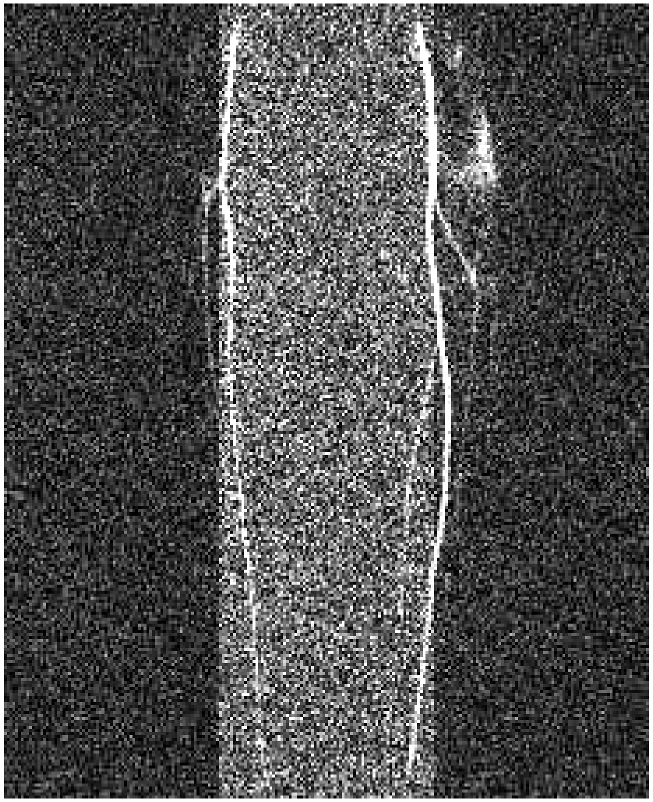
MAP-SENSE Preliminary Results

- Scans acceleraty 5x
- The angiogram was computed by:

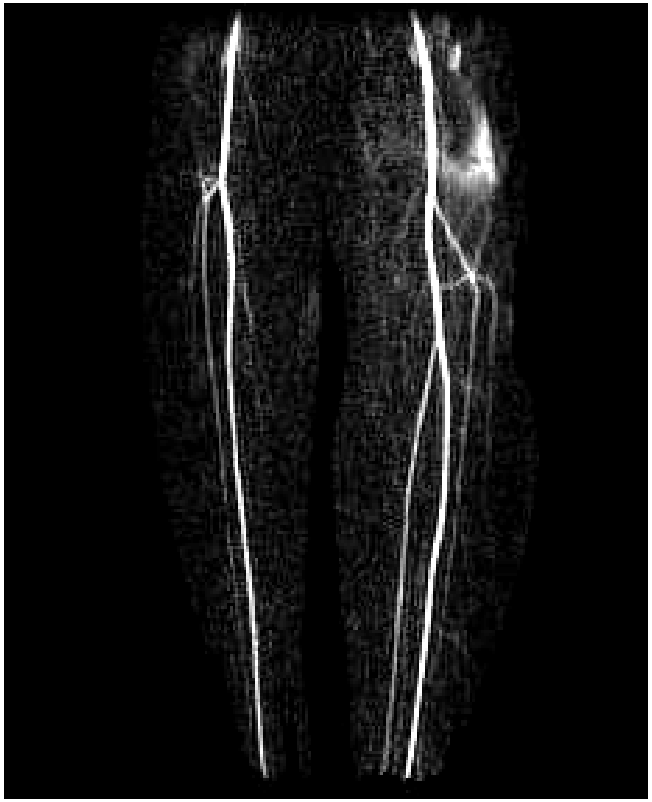
  avg(post-contrast) – avg(pre-contrast)



Unaccelerated



5x faster: SENSE



5x faster: MAP-SENSE

Part V : Time Series Analysis

Finding statistically optimal models for time series data

Time Series Analysis

- Lots of clinical data are time series
  - ECG, MR contrast enhanced tumor data,
  - ER caseload, cancer patient survival curves,…
- How to extract temporal features from these data?
  - Use Fourier, Wavelet Transforms, fit piecewise linear models
  - Good, but… features are arbitrary, no optimality properties
- Now: better models for time series with memory
  - Autoregressive (AR) model
  - Moving Average (MA) model
  - Both AR + MA

Autoregressive models

- When current signal depends on past signal
  - Data has "memory"
- Quite natural for biological and physiological data
  - Because real systems cannot change arbitrarily
  - Current state depends on previous state
- Model:

$$x_{i+1} = \phi_1 x_i + \phi_2 x_{i-1} + \cdots + \phi_p x_{i-p+1} + \xi_{i+1}$$

AR parameters

Noise or input process

How to estimate phi1?

Estimating AR parameters

- Example: For $p = 1 \quad x_{i+1} = \phi_1 x_i + \xi_{i+1}$

$$\underbrace{\begin{pmatrix} x_2 \\ x_3 \\ \vdots \\ x_N \end{pmatrix}}_{\mathbf{b}} = \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix}}_{\mathbf{A}} \phi_1$$

- If noise is independent Gaussian, then show that ML estimate is

$$\hat{\phi}_1 = \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{b}$$

Estimating AR parameters

$$p = 2 \qquad x_{i+1} = \phi_1 x_i + \phi_2 x_{i-1} + \xi_{i+1}$$

$$\underbrace{\begin{pmatrix} x_3 \\ x_4 \\ \vdots \\ x_N \end{pmatrix}}_{\mathbf{b}} = \underbrace{\begin{pmatrix} x_2 & x_1 \\ x_3 & x_2 \\ \vdots & \vdots \\ x_{N-1} & x_{N-2} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}}_{\mathbf{\Phi}}$$

- Can you do this for any p?

More on this at:

http://www-stat.wharton.upenn.edu/~steele/Courses/956/Resource/YWSourceFiles/YW-Eshel.pdf

Moving Average Process

- MA process model:

Model order

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Noise or input process

MA parameters

- Autoregressive moving average (ARMA) model:
- Combines both AR and MA processes

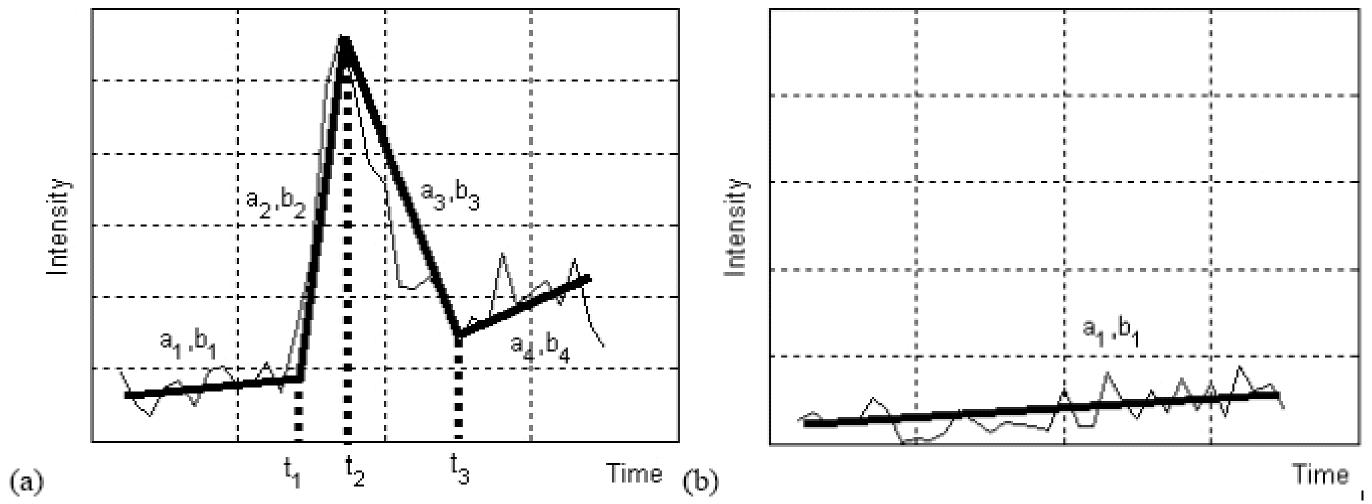
$$X_t = \boxed{c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i}} + \boxed{\sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.}$$

Estimating MA and ARMA parameters

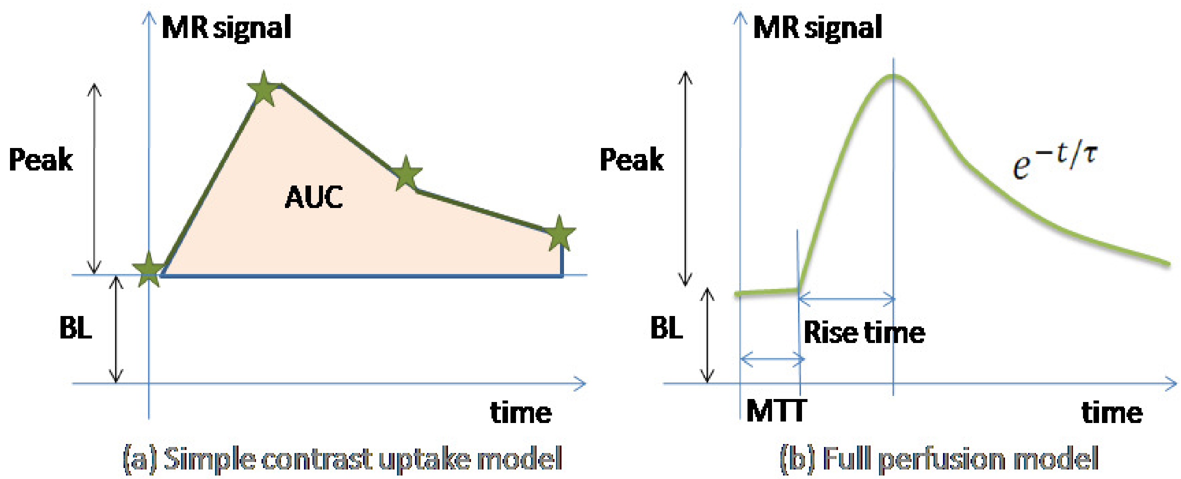
- MA parameters = from Fourier Transforms
- ARMA: more complicated, but most languages have library routines that do this
- E.g. in MATLAB: ar(), arma(), etc
- What is the benefit of estimating AR and ARMA parameters?
  - Might provide better temporal features than FT, WT or arbitrary temporal features
- Example: could apply this to mammography or liver tumor data

Examples

Artery/vein in MR angiography



Liver Tumour time profiles



- Maybe these are better modeled as ARMA???

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

Next lecture (Friday)

- Non-parametric density estimation
  - Histograms + various fitting methods
  - Nearest neighbor
  - Parzen estimation

Next lecture (Wednesday)

- Maximum A Posteriori Estimators
  - Several examples

CS5540: Computational Techniques for Analyzing Clinical Data
Lecture 7:

Statistical Estimation: Least Squares, Maximum Likelihood and Maximum A Posteriori Estimators

Ashish Raj, PhD

Image Data Evaluation and Analytics Laboratory (IDEAL)

Department of Radiology

Weill Cornell Medical College

New York