Image Denoising

Suyash P. Awate

Bayesian Image Denoising

- Optimal noiseless image is the one that maximizes the posterior PDF

$$P(\text{NoiselessImage}|\text{NoisyImage}) = \frac{P(\text{NoisyImage}|\text{NoiselessImage})P(\text{NoiselessImage})}{P(\text{NoisyImage})}$$

  - **Likelihood** PDF = **noise** model = probability of generating the data given the noiseless image
  - **Prior** PDF = our prior beliefs about the noiseless image **before** observing the data
  - **Posterior** PDF: product of likelihood and prior
    - What we get "post" / after observing the data

Optimization for Denoising

- Noiseless image $X = x$
  - X is a MRF
- Observed image data $Y = y$
- Noise model for intensities given noiseless intensities (i.i.d) $P(Y|X) := \Pi_i \; P(Y_i | X_i)$
  - e.g., If noise is additive i.i.d. zero-mean Gaussian, $P(Y_i | X_i) = G(y_i | x_i, \sigma^2)$
- Let θ = parameters underlying noise model and MRF model (in general)

Optimization for Denoising

- Optimization Problem and Strategy

  (1) Assume : MRF parameters are user controlled

  - No need to optimize

  (2) Assume : noise level is already known

  - e.g., using the ML estimate in the background region, where signal is known to be zero

  (3) Get MAP estimate for noiseless image x :

  $\max_x P(x \mid y, \theta)$

Optimization for Denoising

- Lets see what happens at voxel i ?
- Rewrite the objective function
  - P (X | y, theta)

$= P(X_i, X_{\sim i} | y, \theta)$

$= P(X_i | X_{\sim i}, y, \theta) P(X_{\sim i} | y, \theta)$ Conditional Probability

$= P(X_i | X_{N_i}, y, \theta) P(X_{\sim i} | y, \theta)$ Markov assumption on X

$= P(X_i | X_{N_i}, y_i, \theta) P(X_{\sim i} | y, \theta)$ Conditional independence assumption in noise model

Optimization for Denoising

- Optimization Algorithm 1
  - Iterated Conditional Mode (ICM)
  - For a moment, consider optimization over a single voxel i
    - Perform $\max_{x_i} P(X \mid y, \theta)$

$$= \max_{x_i} P(X_i|X_{N_i}, y_i, \theta) P(X_{\sim i}|y, \theta)$$

$$= \max_{x_i} P(X_i|X_{N_i}, y_i, \theta)$$ Second term doesn't depend on x_i

$$= \max_{x_i} \frac{P(y_i|X_i, X_{N_i}, \theta) P(X_i|X_{N_i}, \theta)}{P(y_i|X_{N_i}, \theta)}$$ Bayes Rule

$$= \max_{x_i} P(y_i|X_i, X_{N_i}, \theta) P(X_i|X_{N_i}, \theta)$$ Denominator doesn't depend on x_i

$$= \max_{x_i} P(y_i|X_i, \theta) P(X_i|X_{N_i}, \theta)$$ Conditional independence assumption in noise model

Optimization for Denoising

- Optimization Algorithm 1
  - Iterated Conditional Mode (ICM)
  - $\max_{x_i} P(y_i|X_i,\theta)P(X_i|X_{N_i},\theta)$
  - 1st term P (yi | Xi , θ) = likelihood function
    - Noise model
  - 2nd term P (Xi |X_{Ni} , θ) = local/conditional prior on noiseless image
    - Image-regularity / smoothness model
  - ICM seeks the mode of the local/conditional posterior !

Optimization for Denoising

- Various Optimization Algorithms
  - Order of Intensity Updates:
    - We want every update to increase the posterior P (x|y, θ)

      (1) Sequentially: Column by column, and then row by row
      - May lead to artifacts

      (2) Sequentially: Randomized order each iteration
      - Need to generate random sequence each iteration
      - Are artifacts eliminated ?

      (3) In Parallel: If seeking mode, doesn’t guarantee increase in posterior probability (INVALID)

      (4) In Parallel: Don't go to the mode; go towards the mode
      - Gradient ascent : Dynamic step size + Objective-function monitoring
      - Guarantees increase in posterior probability

- Note on gradient ascent
  - Dynamic step sizing at each iteration :
    - Increase step size by (say) 10% when initial step size increases probability
      - Prevents very slow convergence when far away from optimum
    - Decrease step size by (say) 50% when initial step size decreases probability
      - Adapts step size as you get close to optimum
    - This dual strategy also prevents over-sensitivity to initial step size
  - Termination criteria :
    - Allowable step size becomes very small, e.g., 1e-8
    - Step doesn't increase probability by much, e.g., 0.01% of probability at current solution

Optimization for Denoising

- Various Optimization Algorithms
  - Gradient ascent needs :

(1) Derivative of local conditional PDF w.r.t. xi

$$P(X_i|X_{N_i}, \theta) = \frac{1}{Z_i} \exp\left(-\sum_{a \in A} V_a(x_a)\right)$$

where A is the set of cliques that contain site i

(2) Derivative of noise model w.r.t. xi

MRI (Complex) (Noise: Gaussian)

- Noise Model
  - Circularly-symmetric uni-variate Gaussian

$$P(y|x) = \Pi_i P(y_i|x_i) = \Pi_i G_{\mathbb{C}}(y_i|x_i, \sigma^2) = \Pi_i \frac{1}{\sigma^2 \pi} \exp\left(-\frac{|y_i - x_i|^2}{\sigma^2}\right)$$

MRI (Complex) (Noise: Gaussian)

- For ICM optimization, at chosen voxel i, perform

$$\max_{x_i} P(x|y,\theta) = \max_{x_i} P(y_i|x_i,\theta)P(x_i|x_{N_i},\theta)$$

$$= \max_{x_i} \left(\log P(y_i|x_i,\theta) + \log P(x_i|x_{N_i},\theta) \right)$$

$$= \max_{x_i} \left(\frac{-|y_i - x_i|^2}{\sigma^2} + \sum_{a \in A_i} (-V_a(x_a)) \right)$$

$$= \min_{x_i} \left(\frac{|y_i - x_i|^2}{\sigma^2} + \sum_{a \in A_i} V_a(x_a) \right)$$

  - 1st term = Fidelity term = penalizes the deviation (infidelity) of the estimate x from the data y
  - 2nd term = Regularity term: penalizes roughness of x

MRI (Complex) (Noise: Gaussian)

- For gradient-descent optimization, at voxel i, the derivative is :

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = \frac{1}{\sigma^2} 2(x_i - y_i) + \frac{\partial}{\partial x_i} \sum_{a \in A_i} V_a(x_a)$$

- For the entire image x,
the gradient (column vector) is :

$$g_2(x) := \left(\cdots, \frac{\partial P(x|y,\theta)}{\partial x_i}, \cdots \right)$$

- Current solution x^n at iteration n. Stepsize τ .
Updated solution is : $x^{n+1} = x^n - \tau g(x)$

MRI Magnitude Images (Noise: Rician)

- Observed noisy (magnitude-MR) image data y is real
- Noiseless (magnitude-MR) image x is real
- Prior PDF remains same
- Likelihood PDF is :

$$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i \frac{y_i}{\sigma^2} \exp\left(-\frac{y_i^2 + x_i^2}{2\sigma^2}\right) I_0\left(\frac{y_i x_i}{\sigma^2}\right)$$

where I0 (z) is the modified Bessel function of the first kind with order zero

MRI Magnitude Images (Noise: Rician)

- For ICM optimization, at a chosen voxel i, perform

$$\max_{x_i} P(x|y,\theta) = \min_{x_i} \left(\frac{y_i^2 + x_i^2}{2\sigma^2} - \log I_0 \left(\frac{y_i x_i}{\sigma^2} \right) + \sum_{a \in A_i} V_a(x_a) \right)$$

MRI Magnitude Images (Noise: Rician)

- For gradient-descent optimization, at a chosen voxel i, the derivative is :

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = \frac{x_i}{\sigma^2} - \frac{I_1\left(\frac{y_i x_i}{\sigma^2}\right)}{I_0\left(\frac{y_i x_i}{\sigma^2}\right)}\frac{y_i}{\sigma^2} + \frac{\partial}{\partial x_i}\sum_{a\in A_i} V_a(x_a)$$

where I1 (z) is the modified Bessel function of the first kind with order one

Ultrasound Magnitude Images (Noise: Speckle)

- Observed noisy image data y is real
- Noiseless image x is real
- Prior PDF remains same
- Speckle-Noise model is : $Y = X + \sqrt{X}\, Z$ where $P(Z) := G(0, \sigma^2)$
- What is $P(Y \mid X)$?
  - We need this because this is the likelihood PDF !

Ultrasound Magnitude Images (Noise: Speckle)

- What is P (Y | X) ?
  - Use transformation of random variables

RV $Z, P(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right)$

RV $Y_1 := \sqrt{x}Z, P(y_1) = \frac{1}{\sqrt{x}\sigma\sqrt{2\pi}} \exp\left(-\frac{y_1^2}{2x\sigma^2}\right)$

Transformation maintains PDF as Gaussian and scales std. dev. by $\sqrt{x}$

RV $Y_2 := Y_1 + x, P(y_2) = \frac{1}{\sqrt{x}\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_2 - x)^2}{2x\sigma^2}\right)$

Transformation maintains PDF as Gaussian and translates mean by x

  - So, the likelihood PDF is :

$$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i G(y_i|x_i, x_i\sigma^2) = \frac{1}{\sqrt{x}\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - x_i)^2}{2x_i\sigma^2}\right)$$

Ultrasound Magnitude Images (Noise: Speckle)

- For ICM optimization, at a chosen voxel i, perform

$$\max_{x_i} P(x|y,\theta) = \min_{x_i} \left(\frac{\log(x_i)}{2} + \frac{(y_i - x_i)^2}{2 x_i \sigma^2} + \sum_{a \in A_i} V_a(x_a) \right)$$

Ultrasound Magnitude Images (Noise: Speckle)

- For gradient-descent optimization, at voxel i, the derivative is :

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = \frac{1}{2x_i} + \frac{1}{2\sigma^2}\left(\frac{x_i^2 - y_i^2}{x_i^2}\right) + \frac{\partial}{\partial x_i}\sum_{a \in A_i} V_a(x_a)$$

Weighting Likelihood & MRF Prior

- Motivation

  (1) We want to control the strength of the prior model based on certain criteria (e.g., noise level)

  (2) We want to balance the enforcement of fidelity and regularity based on certain criteria (e.g., noise level)

- Thumb Rules

  (1) Very high noise levels ⇒ data is highly corrupted ⇒ use a strong prior

  (2) Very low noise levels ⇒ data is high quality ⇒ use a weak prior

Weighting Likelihood & MRF Prior

- How to do it ?
  - Introduce a user-controlled parameter $\beta \in [0, 1]$ that defines the balance between :
    - (1) enforcing the prior model and
    - (2) enforcing the likelihood model

Weighting Likelihood & MRF Prior

- Weighting MRF Prior
  - In the prior PDF, introduce a parameter β ∈ [0, 1] s.t.

$$P(x) := \frac{1}{Z(\beta)} \exp\left(-\beta \frac{1}{T} U(x)\right) \text{ where}$$

$$U(x) := \sum_{c \in C} V_c(x_c) \text{ where}$$

$$Z(\beta) := \sum_{x} \exp\left(-\beta \frac{1}{T} U(x)\right)$$

  - This changes the local conditional prior to

$$P(x_i | x_{N_i}, \theta) = \frac{1}{Z_i(\beta)} \exp\left(-\beta \frac{1}{T} \sum_{a \in A} V_a(x_a)\right)$$

  - Introducing β is similar to changing the temperature T

Weighting Likelihood & MRF Prior

- Weighting Likelihood (Complex-Gaussian Noise)
  - Introduce a parameter α ∈ [0, 1], where α := 1 − β
  - $$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i G_\alpha(y_i|x_i, \sigma^2)$$
    $$G_\alpha(y_i|x_i, \sigma^2) := \frac{1}{Z(\sigma, \alpha)} \exp\left(-\alpha \frac{|y_i - x_i|^2}{\sigma^2}\right),$$
    where $Z(\sigma, \alpha) = 1 / ((\sigma/\alpha)^2 \pi)$
  - Interpretation
    - Introducing α is similar to changing the noise level / standard deviation σ
    - This is the Complex-Gaussian PDF with parameters (xi , $\sigma^{2/}\alpha$)

Weighting Likelihood & MRF Prior

- Modified Optimization Problem (Complex-Gaussian Noise)
  - For ICM optimization, at a chosen voxel i, perform :

$$\max_{x_i} P(x|y,\theta) = \min_{x_i} \left((1-\beta)\frac{|y_i - x_i|^2}{\sigma^2} + \beta \sum_{a \in A_i} V_a(x_a) \right)$$

  - For gradient-descent optimization, at a chosen voxel i, the derivative is :

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = (1-\beta)\frac{1}{\sigma^2} 2(x_i - y_i) + \beta \frac{\partial}{\partial x_i} \sum_{a \in A_i} V_a(x_a)$$

  - β=0 (T→∞) ignores the prior; we get the ML estimate
  - β=1 (α = 0; σ→∞) makes the likelihood a uniform PDF

Weighting Likelihood & MRF Prior

- Weighting Likelihood (Rician Noise)
  - Introduce a parameter $\alpha \in [0, 1]$, where $\alpha := 1 - \beta$

$$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i \frac{\alpha y_i}{\sigma^2} \exp\left(-\alpha \frac{y_i^2 + x_i^2}{2\sigma^2}\right) I_0\left(\alpha \frac{y_i x_i}{\sigma^2}\right)$$

  - Interpretation
    - Introducing α is similar to changing the noise level σ
    - This is the Rician PDF with parameters (x_i , σ / sqrt(α))

Weighting Likelihood & MRF Prior

- Modified Optimization Problem (Rician Noise)
  - For ICM optimization, at a chosen voxel i, perform

$$\max_{x_i} P(x|y,\theta) = \min_{x_i} \left((1-\beta)\frac{y_i^2 + x_i^2}{2\sigma^2} - \log I_0\left((1-\beta)\frac{y_i x_i}{\sigma^2}\right) + \beta \sum_{a\in A_i} V_a(x_a) \right)$$

  - For gradient-descent optimization, at a chosen voxel i, the derivative is

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = (1-\beta)\frac{x_i}{\sigma^2} - \frac{I_1\left((1-\beta)\frac{y_i x_i}{\sigma^2}\right)}{I_0\left((1-\beta)\frac{y_i x_i}{\sigma^2}\right)}(1-\beta)\frac{y_i}{\sigma^2} + \beta\frac{\partial}{\partial x_i}\sum_{a\in A_i} V_a(x_a)$$

  - β=0 ($T \to \infty$) ignores the prior; we get the ML estimate
  - β=1 (α = 0; $\sigma \to \infty$) makes the likelihood a uniform PDF

Weighting Likelihood & MRF Prior

- Weighting Likelihood (Speckle Noise)
  - Introduce a parameter $\alpha \in [0, 1]$, where $\alpha := 1 - \beta$

$$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i G(y_i|x_i, x_i\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}\sqrt{x_i^\alpha/\alpha}} \exp\left(-\frac{(y_i - x_i)^2}{2(x_i^\alpha/\alpha)\sigma^2}\right)$$

  - Interpretation
    - Introducing α is similar to changing the noise level / standard deviation σ
    - This is the Gaussian PDF with parameters $(x_i, \sigma^2 (x_i^\alpha/\alpha))$

Weighting Likelihood & MRF Prior

- Modified Optimization Problem (Speckle Noise)
  - For ICM optimization, at a chosen voxel i, perform

$$\max_{x_i} P(x|y,\theta) = \min_{x_i} \left(\alpha \frac{1}{2} \log(x_i) + \alpha \frac{1}{2\sigma^2} \frac{(y_i - x_i)^2}{x_i^\alpha} + \beta \sum_{a \in A_i} V_a(x_a) \right)$$

  - For gradient-descent optimization, at a chosen voxel i, the derivative is

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = \alpha \frac{1}{2x_i} + \alpha \frac{1}{2\sigma^2} \frac{\partial}{\partial x_i} \frac{(y_i - x_i)^2}{x_i^\alpha} + \beta \frac{\partial}{\partial x_i} \sum_{a \in A_i} V_a(x_a)$$

  - $\beta=0$ ($T \rightarrow \infty$) ignores the prior; we get the ML estimate
  - $\beta=1$ ($\alpha = 0$; $\sigma \rightarrow \infty$) makes the likelihood a uniform PDF

Possible Project Topics

- A topic of your choice
  - With approval of instructor
- CT Reconstruction (2D)
  - Modeling as a linear system
  - Reconstruction by matrix inversion
  - Radon-transform-matrix analysis (SVD, null-space)
- CT Reconstruction (3D)
  - Modeling as a linear system
  - Iterative MAP estimation using Poisson noise model
  - MRF in 3D

Possible Project Topics

- Reconstruction of MRI
  - Sparse image representation
  - Iteratively re-weighted least squares (IRLS)
    - http://math.lanl.gov/Research/Publications/Docs/chartrand-2008-iteratively.pdf
- Reconstruction of Dynamic MRI (space + time)
  - Choice of image-acquisition (Fourier sampling) strategies
  - Sliding-window reconstruction algorithm
    - http://www.ncbi.nlm.nih.gov/pubmed/11870913
  - Reconstruction via Bayesian estimation
- Reconstruction of Diffusion-Tensor MRI
  - Multi-dimensional MRI
  - Fidelity + MRF-smoothness + Tensor-model fit

Possible Project Topics

- Denoising using patch-based statistics
  - Non-local-means algorithm, Gaussian noise
    - http://en.wikipedia.org/wiki/Non-local_means
  - Adapting for speckle noise
    - http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4982678&tag=1
- Denoising DTI
  - Rician noise model + MRF-smoothness + Tensor-model fit
- Denoising MRI / CT
  - Soft-thresholding algorithm
  - Using wavelet transform for image representation
  - Analysis performance with different wavelets
  - http://eeweb.poly.edu/iselesni/WaveletSoftware/denoise.html

Possible Project Topics

- Clustering of time series in functional-MRI (space + time)
  - Time-series data at each voxel
  - Modeling distributions of normalized (zero mean, unit variance) time-series on hypersphere
  - Estimation
  - http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3513676/
- Segmentation of multimodal brain MRI
  - Denoise T1 + T2 MRI images, of same person, jointly
  - Extend 1D Gaussians to multi-D Gaussians
  - Model Gaussian covariances
  - Optimization via EM
- Fuzzy-C-means based brain-segmentation algorithms
  - http://link.springer.com/article/10.1007%2Fs10462-012-9318-2

Possible Project Topics

- Shape analysis of anatomical structure (2D)
  - e.g., an anatomical structure in brain
- Algorithms for sampling images from a MRF model
  - Gibbs sampling for label images
  - Texture sampling for MR / CT images
- Algorithms for sampling shapes from a shape model
  - Hybrid Monte Carlo

Grading Policy - Minor Update

- Mid-sem exam – 10%
- End-sem exam – 20%
  - Will be simple; may be open book
- Course project – 20% (due towards Apr end)
- Assignments – 4 x 10% = 40%
- Quiz (in April) – 5%
- Class Participation – 5%