Image Processing with Nonparametric Neighborhood Statistics II

Ross T. Whitaker

Scientific Computing and Imaging Institute
School of Computing
University of Utah

Variational Methods
E.g Anisotropic Diffusion

- Perona&Malik (1990)

$$\frac{\partial f}{\partial t} = \nabla \cdot c(|\nabla f|)\nabla f$$

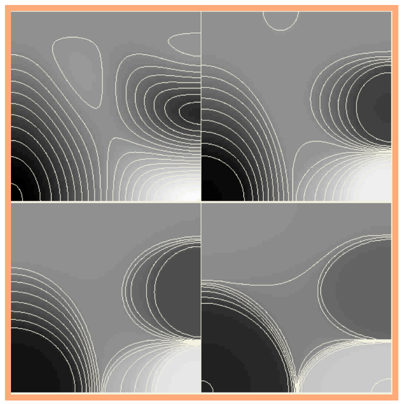
- Penalty:
  - Quadratic on grad-mag with outliers (discontinuities)
    - Nordstrom 1990; Black et. al 1998
  - Favors piecewise const. Images

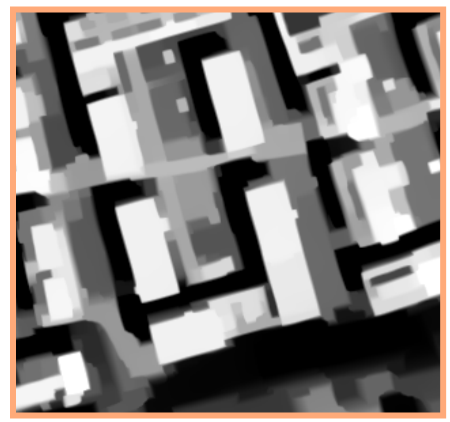
Other Flattening Approaches

- Total variation
  - Rudin et. al (1992)
- Mumford-Shah (1989) related
  - Explicit model of edges
  - Cartoon model
- Level sets to model edges
  - Chan & Vese (2000)
  - Tsai, Yezzi, Willsky (2000)
- Model textures + boundaries
  - Meyer (2000)
  - Vese & Osher (2002)

PDE Methods
Other Examples

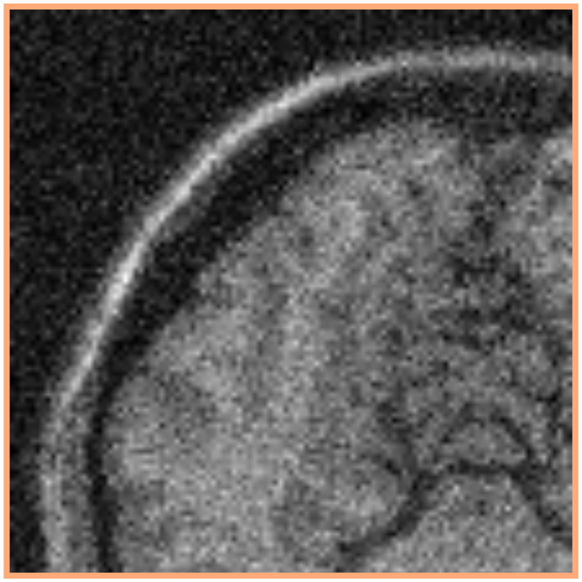
- Weickert (1998)
  - Coherence enhancing
- Tasdizen et. al (2001)
  - Piecewise-flat normals
- Droske et. al (2004)
  - Minimize curvature
- Berkels et. al (2006)
  - Wolf shapes



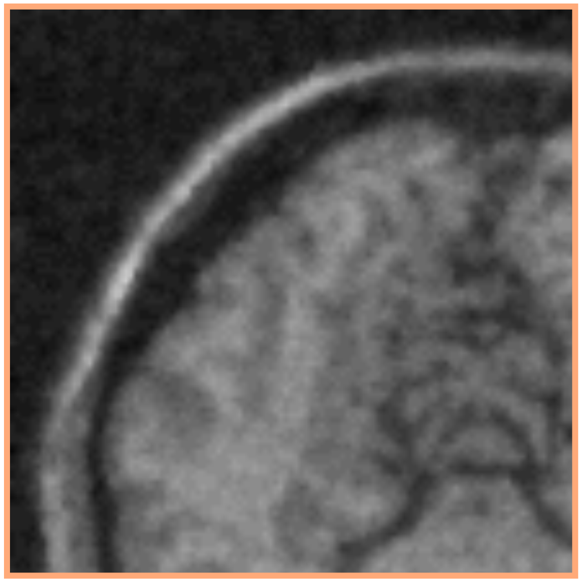


Examples

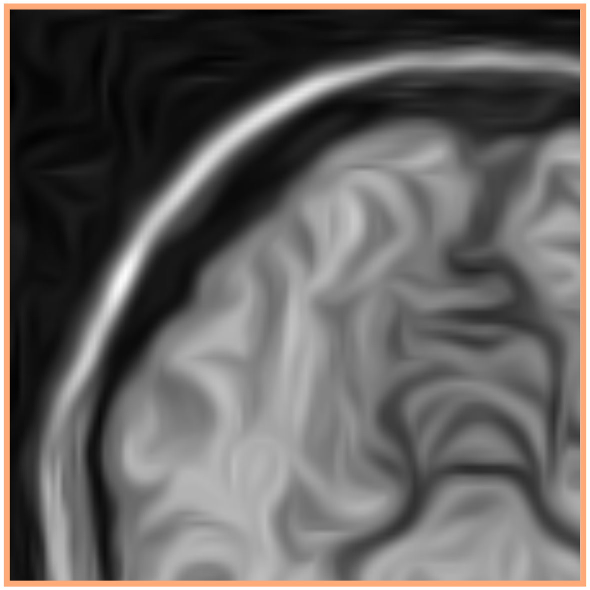
MRI (Simulated noise)



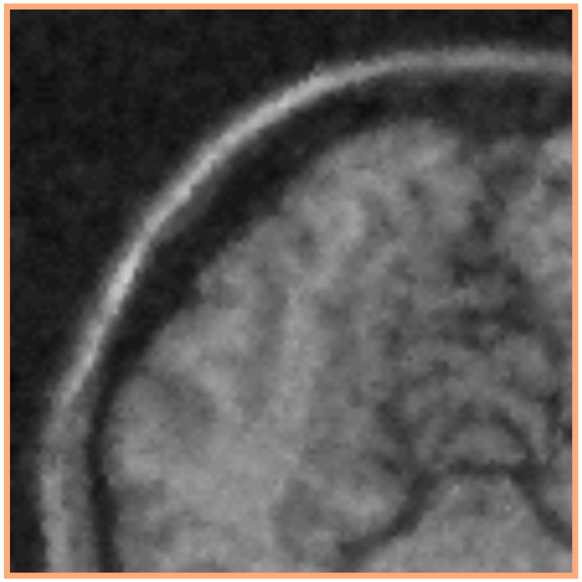
Bilateral Filtering



Coherence Enhancing



Anisotropic Diffusion



Issues

- Prioritize geometric configurations a priori
  - Works well of the model fits, otherwise…
- Free parameters
  - Thresholds -> determine when to apply different models (e.g. “preserve edge or smooth”)
- Generality
  - Cartoon-like simplifications are disastrous in many applications
- Increasing the geometric complexity
  - Is there a better way?

Proposed Strategy

- Infer the appropriate Markovian relationships from the data
- Images neighborhoods as random processes
- Move away from geometric formulations

Nonparametric, Multivariate Density Estimation

- Nonparametric estimation
  - No prior knowledge of densities
  - Can model *real* densities
- Statistics in higher dimensions
  - Curse of dimensionality (volume of n-sphere -> 0)
  - \+ However, empirically more optimistic
  - \+ Z has identical marginal distributions
  - \+ Lower dimensional manifolds in feature space

Nonparametric Density Estimation

- How to estimate a density function f(x) ?
  - Choose parametric form p(Z, θ), and estimate θ
- Suppose we don't have a parametric form
  - BUT we have are lots of samples (n) from the stochastic process we wish to model

Kernel function

$$\phi(u), \text{ s.t. } \phi(u) \geq 0, \int \phi(u)du = 1.$$

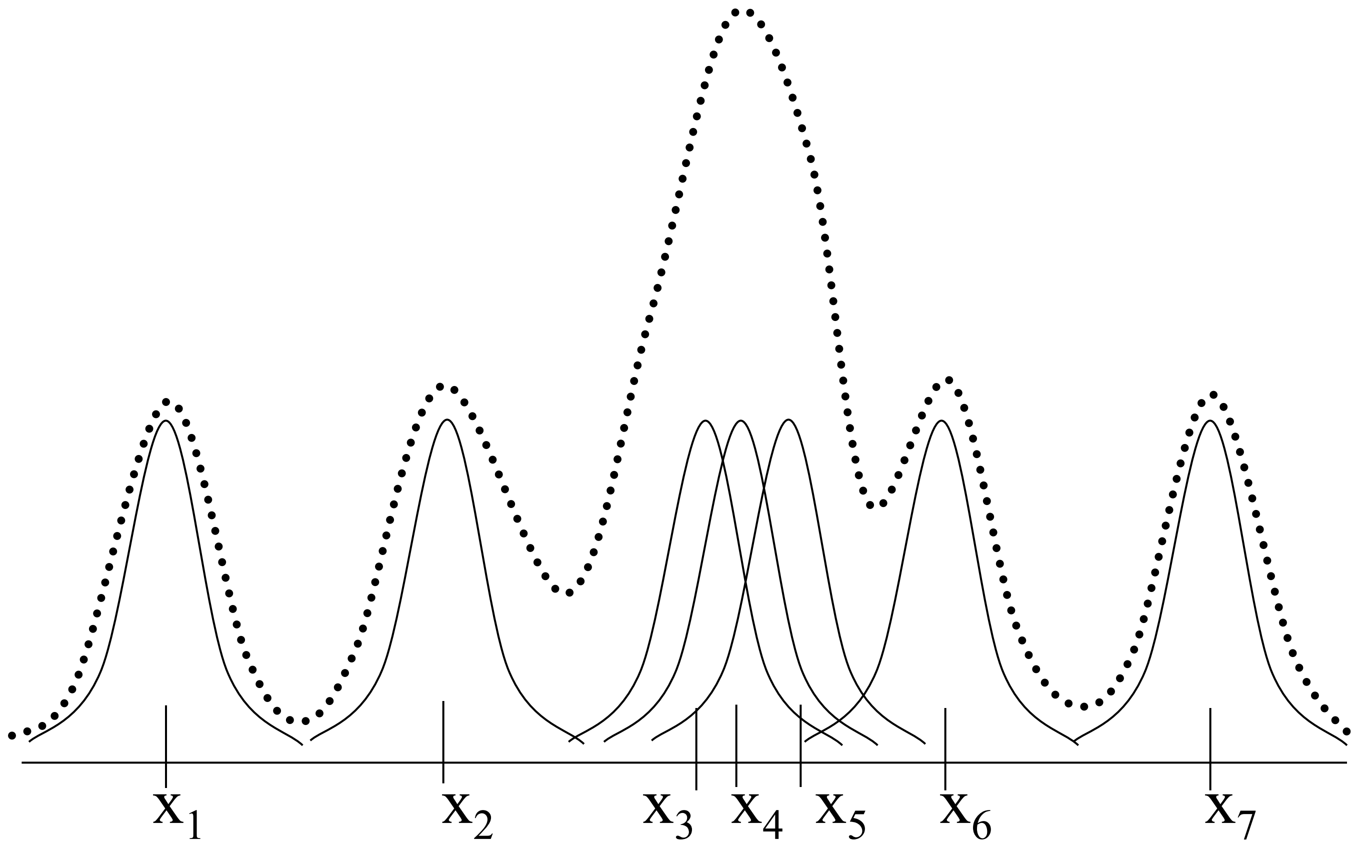
E.g. Gaussian or unit cube

Density estimate

$$p_n(x) = (1/n)\sum_{i=1}^{n} \frac{1}{V_n}\phi(\frac{x - x_i}{h_n}).$$

Normalization $V_n = h_n^d.$

Nonparametric Density Estimation



Bandwidth as a Function of n

- Goal (convergence)

$$p_n(x) \mapsto f(x) \text{ as } n \mapsto \infty.$$

- Conditions

$$\begin{aligned} \lim_{n\to\infty} h_n^d &= 0, \text{ and} \\ \lim_{n\to\infty} n h_n^d &= \infty. \end{aligned}$$

- E.g.

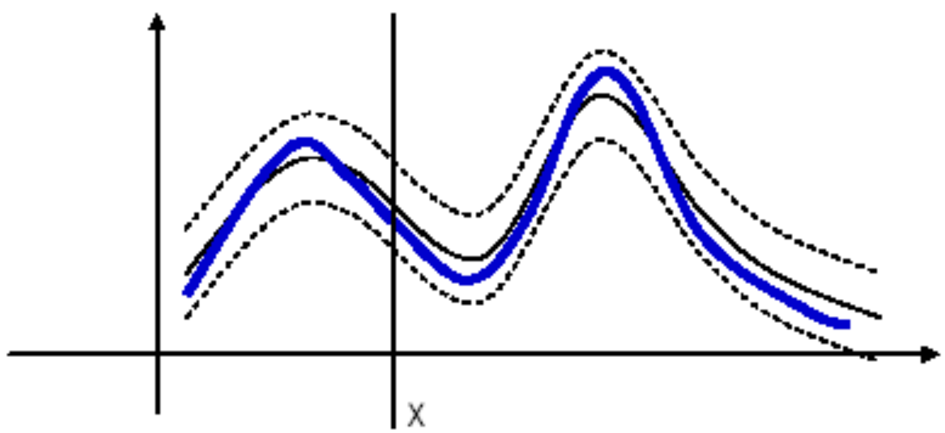
$$h_n^d = 1/\sqrt{n}$$

- Can also fix number of samples in window
  - E.g. KNN

Nonparametric Density Converges

- Show that Parzen window estimator converges to the true density pointwise with increasing number of samples.

$$\lim_{n \mapsto \infty} p_n(x) \mapsto f(x).$$



At each point x, we prove the convergence of mean and the variance decreases to zero

Convergence of Mean

- Take expectation w.r.t. samples from f().

Let $\hat{p}_n(x) = E\{p_n(x)\} =$
$(1/n)\sum_{i=1}^{n} E\{(1/V_n)\phi(\frac{x-x_i}{h_n})\}$
$= \int dy f(y)(1/V_n)\phi(\frac{x-y}{h_n})$.

As $n \mapsto \infty$, $(1/V_n)\phi(\frac{x-y}{h_n}) \mapsto \delta(x-y)$

- So if follows

$$\lim_{n\mapsto\infty} E\{p_n(x)\} = f(x)$$

Convergence of Variance

- Take expectation w.r.t. samples from f().

$$\sigma_n^2(x) = \sum_{i=1}^n E\{(\frac{1}{nV_n}\phi(\frac{x-x_i}{h_n}) - \frac{1}{n}\widehat{p}_n(x))^2\}$$

$$= (1/n)\{E_f[(1/V_n^2)\phi^2(\frac{x-x_i}{h_n}) - E_f^2[p_n(x)]\}$$

$$= (1/nV_n)\int(1/V_n)\phi^2(\frac{x-y}{h_n})f(y)dy$$

$$- (1/n)E_f^2[p_n(x)]\}$$

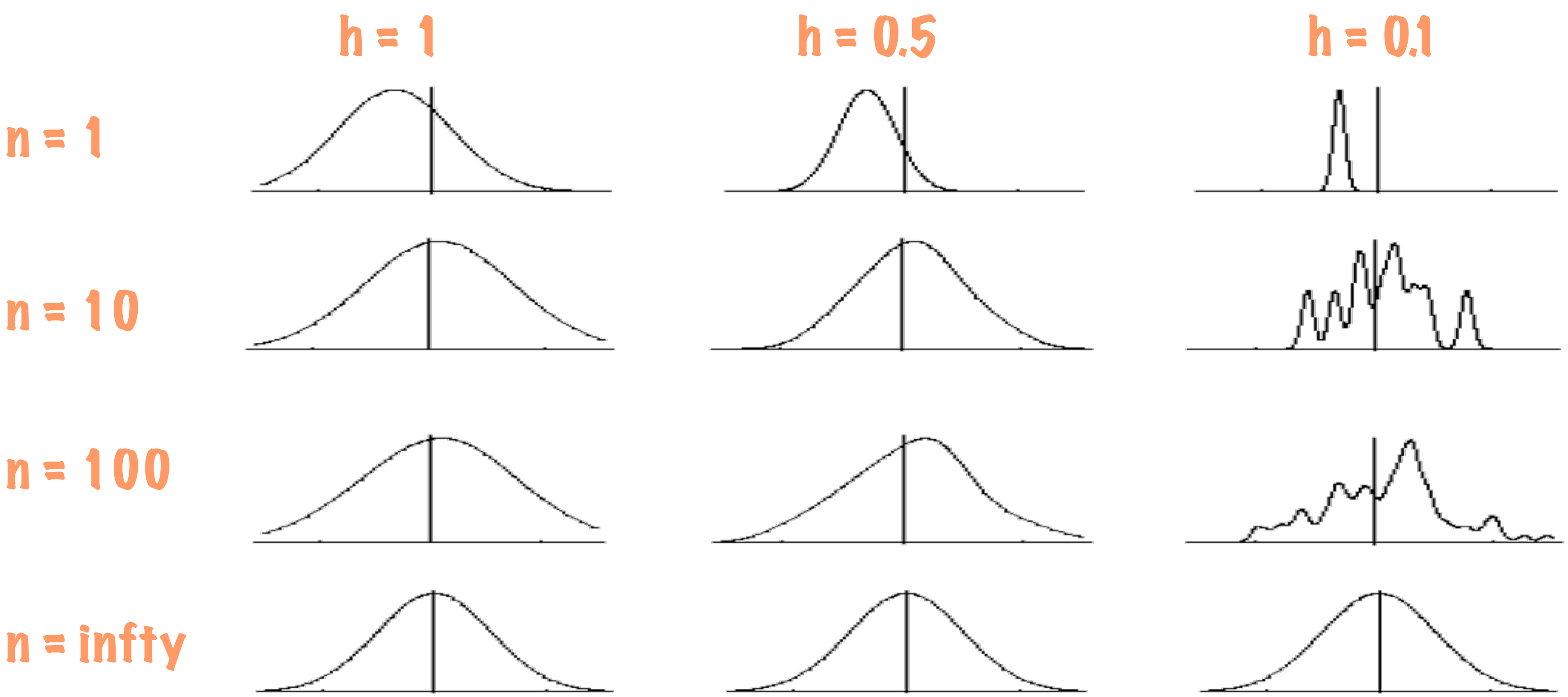
$$\leq \frac{\sup(\phi(.))\widehat{p}_n(x)}{nV_n} \mapsto 0, \quad n \mapsto \infty.$$

- Where

$$\widehat{p}_n(x) = \int \frac{1}{V_n}\phi(\frac{x-y}{h_n})f(y)dy$$

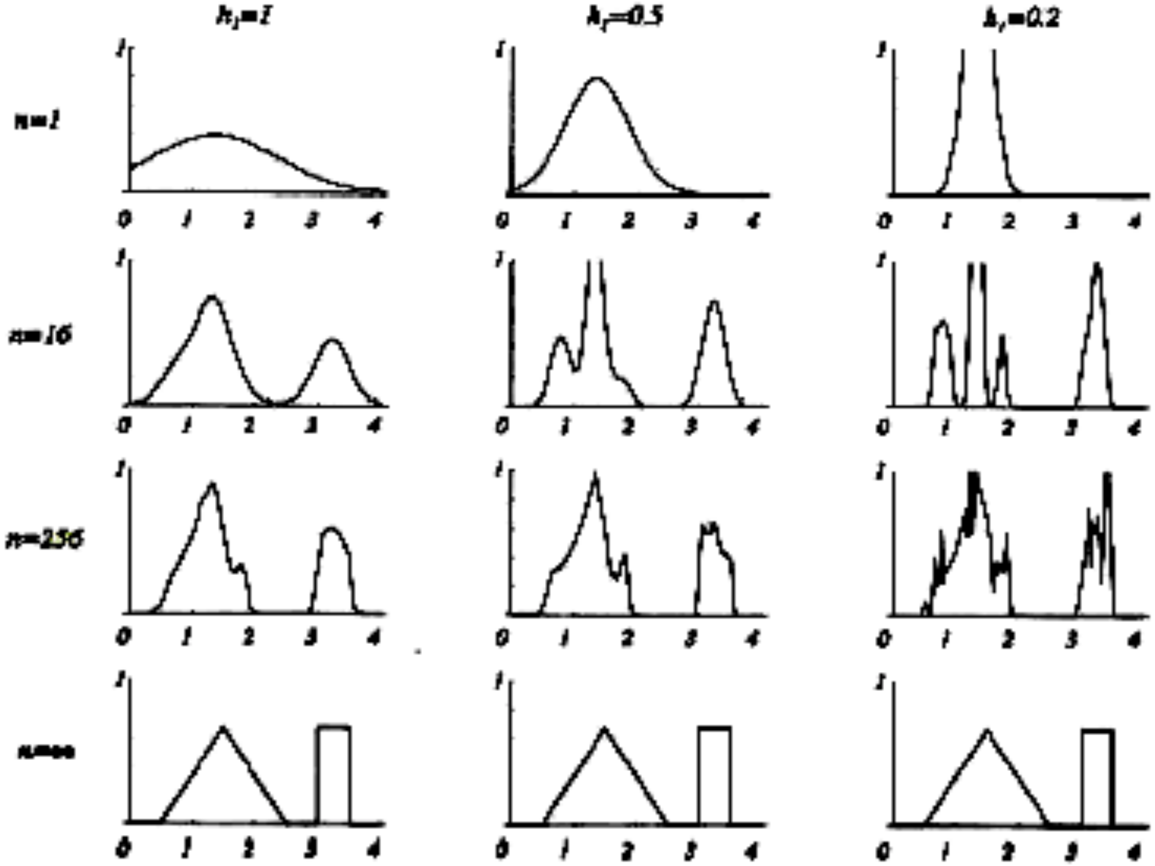
Parzen Windowing Parameters

- Effects of finite sampling (Duda & Hart)



Parzen Windowing Parameters

- Effects of finite sampling (Duda & Hart)



Bias and Variance (Asymptotic)

- Bias increases with kernel size (band width)

$$\mathrm{Bias}(x) = \frac{1}{2} h_n^2 f''(x) \int y^2 \phi(y) dy + \mathrm{o}(h_n^2)$$

- Variance decreases with kernel size

$$\mathrm{Var}(p_n(x)) = (nh_n)^{-1} f(x) \int \phi^2(y) dy + \mathrm{o}((nh_n)^{-1})$$

SCI INSTITUTE

Band Width Selection Problem

- Choosing h for any particular n
- Strategies
  - Analytical

$$h_{\text{opt}}(x) = n^{-1/5} \left(\frac{f(x) \int \phi^2(y) dy}{f''(x) \int y^2 \phi(y) dy} \right)^{1/5}$$

  - Cross validation
    - Maximize the log likelihood of each sample (computed from kernel estimate of all others)

Nonparametric, Multivariate Density Estimation

- Statistics in higher dimensions
  - Curse of dimensionality (volume of n-sphere -> 0)
  - \+ However, empirically more optimistic
  - \+ Z has identical marginal distributions
  - \+ Lower dimensional manifolds in feature space

SCI INSTITUTE

Regression

- Conditional/marginal distribution

$$P(X|Y) = \frac{P(X,Y)}{P_Y(Y)} \qquad P_Y(Y) = \int P(x,Y)dx$$

- Conditional expectation

$$E[X|Y] = \frac{\int xP(x,Y)dx}{P_Y(Y)} = g(y)$$

- Conditional expectation is the "best" funtional model of x=g(y).
  - Why?

SCI INSTITUTE

Nadaraya-Watson Kernel Regression

- Define density estimator on Z=X,Y
  - Separable kernel

$$\phi(z - z_i) = \phi(x - x_i)\phi(y - y_i)$$

  - Use kernel estimates for joint and marginal densities in E[X|Y]

$$\hat{g}(y) = \frac{\sum_i \phi(y - y_i) x_i}{\sum_i \phi(y - y_i)}$$

  - Converges pointwise as n->infty to g(y) = E[X|Y]

MRF Fields and Learning

- Define fields (clique potentials) and parameters a priori
- Define functional form of cliques and estimate parameters from images
- Nonparametric modeling of probability densities of neighborhoods

SCI INSTITUTE

Nonparametric Learning MRF

- Parzen window results require independent samples
- MRF conditions couple pixels in the image
- Mixing property combined with widespread samples ensures sufficient independence
  - Levina 1998

Constructing a Nonparametric Denoising Algorithm

- Assume an image is MRF (+ stationary+ ergodic + very large)
- Suppose we treat each pixel as an unknown w.r.t. it's neighborhood (known)
- What is the best function of it's neighbors we could choose to replace each pixel?

SCI INSTITUTE

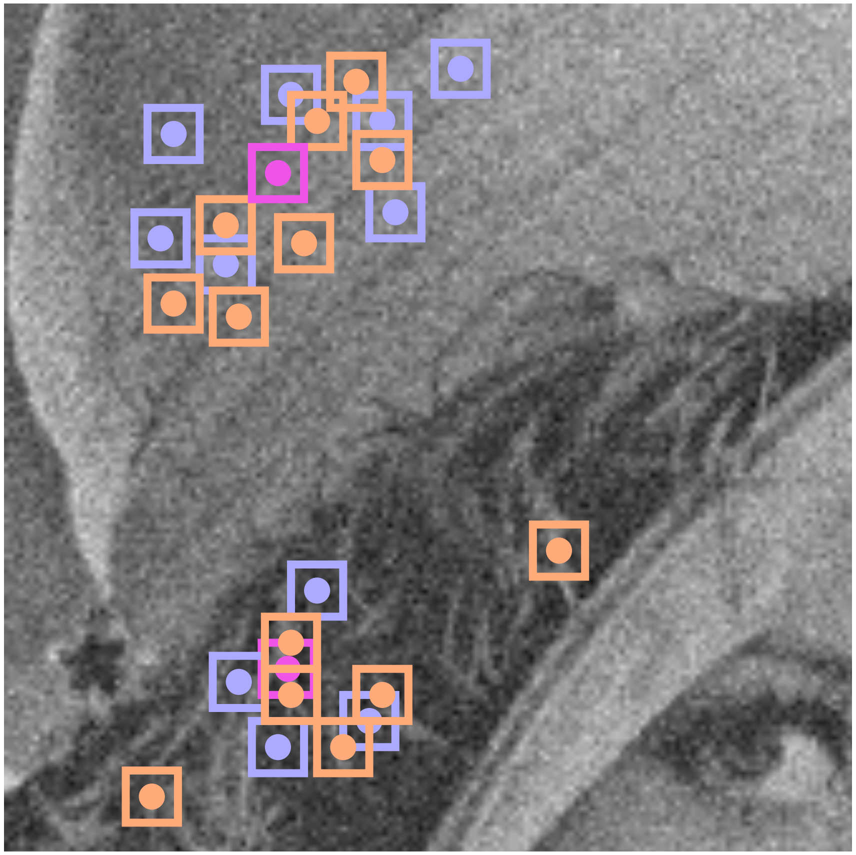
NL-Means
Baudes, Coll, Morel 2005

- Use nonparametric estimate of E[X|Y] for image neighborhoods
  - Replace each pixel in the image with this estimate
- One pass
- Optimal: best function of neighborhood values (minimizes expected squared error)

NL-Means: Details

- In practice, you include the input pixel in the neighborhood comparisons
- Choose samples within MxM block around pixel
  - Nonstationarity, computational efficiency
- Set bandwidth to be proportional to noise
  - CDF is corrupted by noise
  - Limited resolution of recovery

Example: Weights



Example: Results

Bilateral filter (3.0, 0.1)

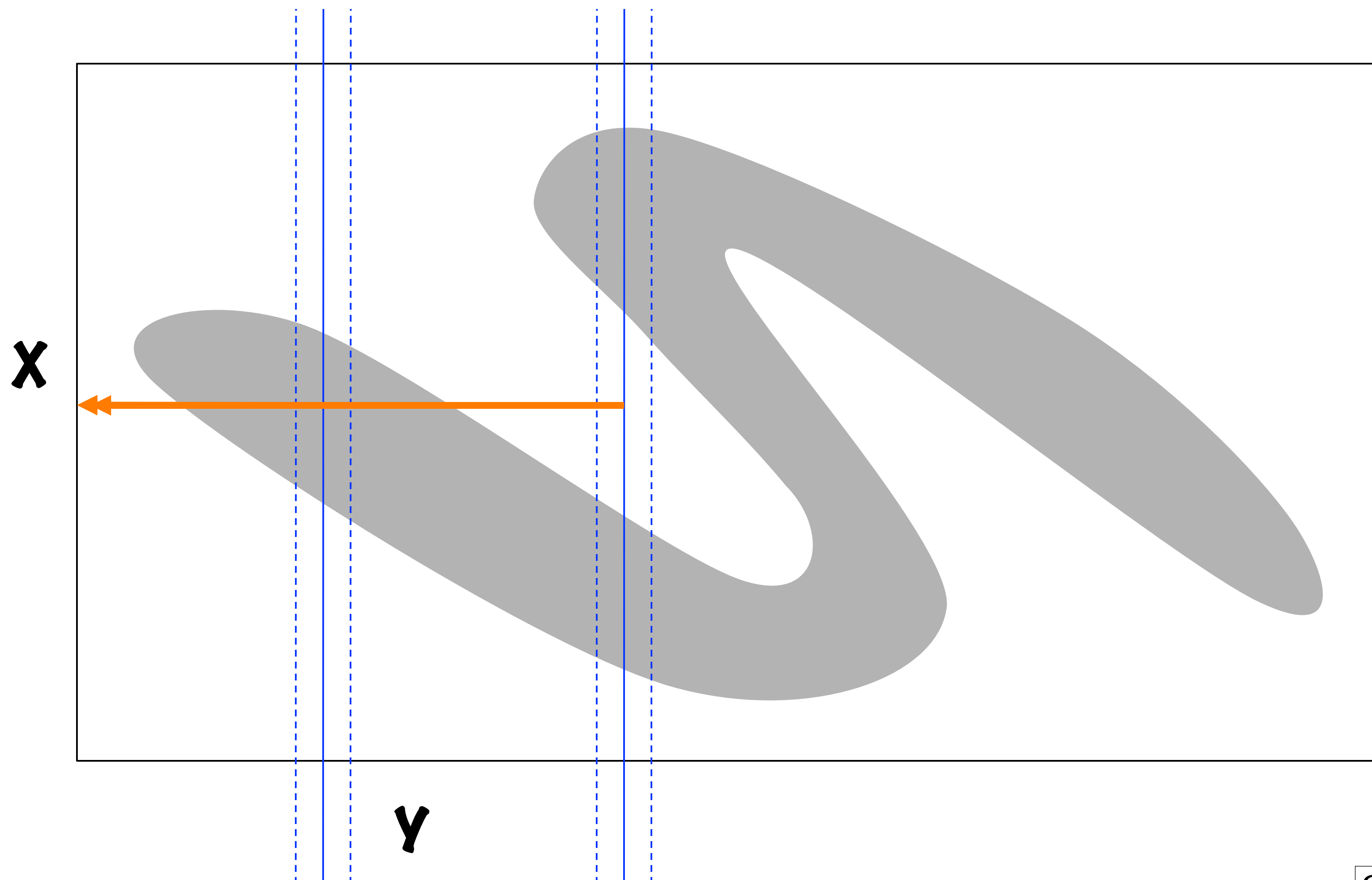
NL means (7, 31, 1.0)

Issues

- Center pixel not part of CDF
  - But it is very helpful in practice
- The conditional expectation is optimal if:
  - You have a good estimate of the MRF
  - You have correct values for neighbors
- In practice:
  - People use the value of the pixel in it's estimate
  - People iterate on this algorithm in one form or another

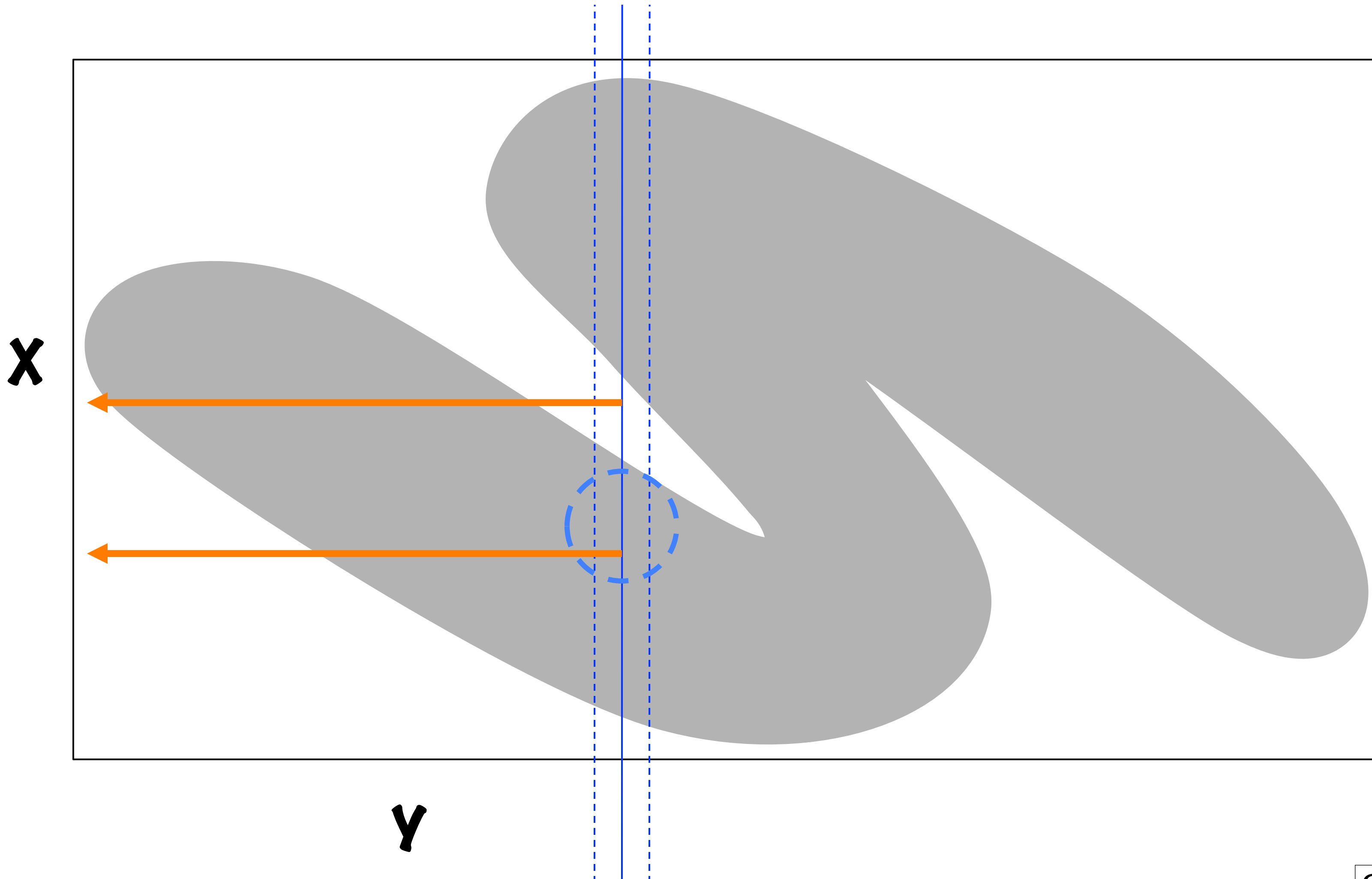
Visualizing NL-Means

Neighborhood space and P(X,Y)



Visualizing NL-Means

Neighborhood space and P(X,Y)



Alternatives Using Posterior

- Construct a conditional Bayesian energy using a nonparametric MRF representation
- Iterated conditional modes (Besag 1974)
  - Converges to local mode of posterior
- Iterated conditional expectation (Owen 1986)
  - Converges to mean field approximation of MRF

- Families of nonlocal averaging algorithms that include noise models, etc.

Other Possibilities

- When iterating are the weights fixed or not?

$$f_I^{k+1} = \frac{\sum_{J \in A} w_{I,J} f_J^k + \lambda(\sum_{J \in A} w_{I,J}) g_I}{(1+\lambda)\sum_{J \in A} w_{I,J}} \qquad W_{I,J} = \exp\left(-\frac{1}{2\sigma^2}\sum_{Q \in \mathcal{N}_0} (f_{I+Q}^k - f_{J+Q}^k)^2\right)$$

- If fixed: Diffusion on a graph in high-dimensional space of neighborhoods
  - Szlam, Maggioni, Coifman 2008
  - Singer, Shkolnisky, Nadler 2009

Break…