

Stat 992: Lecture 37

Anisotropic kernel smoothing II.

Moo K. Chung mchung@stat.wisc.edu

April 28, 2004

1. *Image gradient.* In Lecture 34, we smoothed data $Y(x)$ via iterated kernel smoothing of form

$$K_H * K_H * \cdots * K_H * Y(x)$$

where H is determined by matching it to the Riemannian metric tensor G of data, i.e. $H = 2tG$. The parameter t measures the over all smoothness and by leaving it this way, we are leaving a room for choosing it automatically as a optimal filtering problem.

In diffusion tensor imaging (DTI), the diffusion coefficient gives the the natural metric. We generalize this idea to gray scale images where the metric tensor is not given. The problem is figuring out the most natural metric tensor of the image. Suppose image I can be modelled additively as

$$I(x) = \mu(x) + K_M * W(x)$$

where W is white noise such that $\text{Var}(K_M * W) = 1$. Then it is shown that the smoothness of image can be measured by the variance of the image derivatives, i.e. $\text{Var}(\partial I) = (MM')^{-1}/2$. So this gives a motivation for using the image gradient as the image metric that measures. Unfortunately the image gradient is not well defined. To make it numerically well defined, we differentiate after smoothing. This is the usual way in computer vision area. Also see Fan's nonparametric book where the derivative is estimated this way. Thomas Hoffmann estimated higher order derivatives after spline smoothing. See <http://www.stat.wisc.edu/mchung/papers/HBM2004/HBM2004thomas.html>. Yuefeng Lu is working on derivative estimation problem using polynomial regression. So we estimate the image gradient as

$$\widehat{\partial I} = \partial(K_\sigma * I) = (\partial K_\sigma) * I.$$

Figure 2 left image shows the image gradient estimated this way. Now if we take this as the image metric and match it proportionally to the covariance

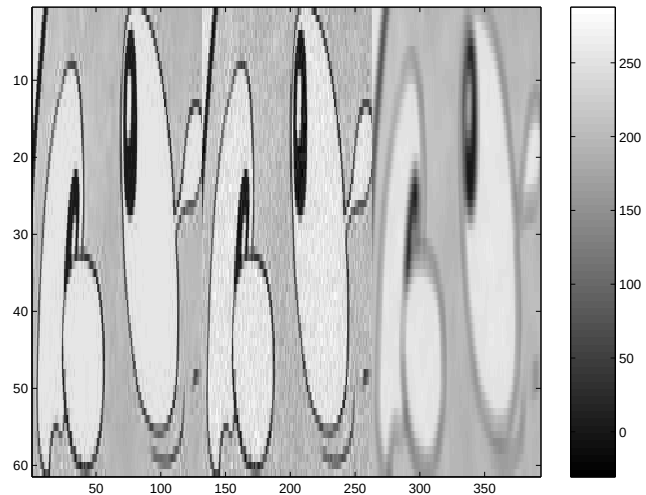


Figure 1: Left: a cartoon image, intensity scale is between 0 and 255. Middle: Gaussian white noise $N(0, 10)$ added. Right: Isotropic kernel smoothing. Noise has been removed inside the big eyes but the boundary has been also blurred.

matrix of kernel, i.e. $HH' = 2tG = 2t\widehat{\partial I}$, metric will be larger near edges so the kernel smoothing will smooth more across the edges which is not what we want to do.

```
file='dooly.jpg'; I=imread(file);
dooly=I(:,:,1);
[xdim, ydim]=size(dooly)
ndooly=dooly+normrnd(0,10,xdim,ydim);
sndooly=gaussblur(ndooly,9);
[vx,vy]=gradient(sndooly);
```

2. *Image metric.* What we need is a smaller metric near the edges. We can construct it in by inverting the gradient but to make it numerically stable, we let

$$G = \begin{pmatrix} 1 + |\widehat{\partial_{x_1} I}| & 0 \\ 0 & 1 + |\widehat{\partial_{x_2} I}| \end{pmatrix}^{-1}.$$

We have taken the absolute value of the gradient to avoid singularity when $\widehat{\partial_{x_i} I} + 1 = 0$. This is fairly

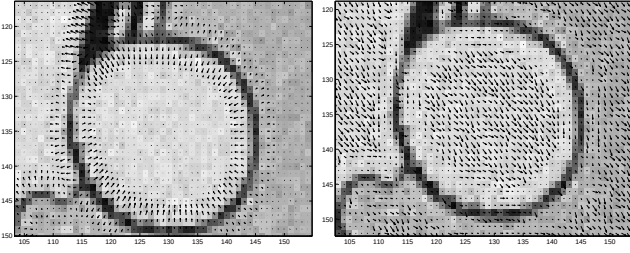


Figure 2: Left: Gradient fields $V = (\partial K_\sigma) * I$. If we use this as the metric, we will smooth more across the edges. We need to come up with the vector fields that will not smooth near edges. Right: fields $1/(1 + |V_i|)$. These vector fields are not smooth enough.

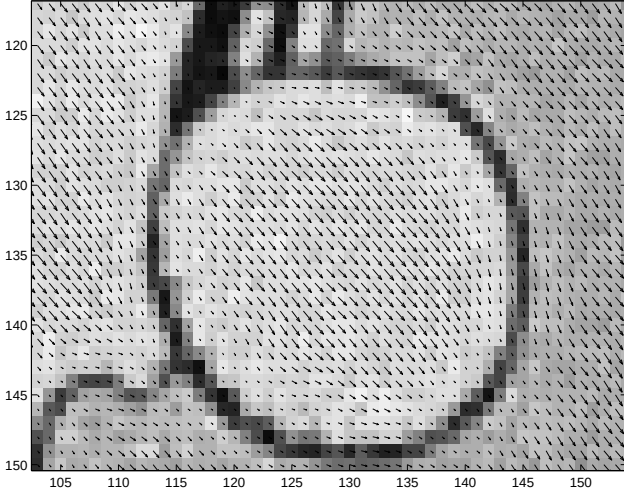


Figure 3: Smoothed vector fields $K_\sigma * 1/(1 + |V_i|)$.

reasonable metric. Similar numerical technique has been also used in Sethian's level set method. There are different kind of metric to investigate. Hessian matrix of the image intensity, i.e. $\partial_{x_i}\partial_{x_j}I$ or the Fisher information matrix, local sample covariance matrix. Unfortunately the above metric will not be smooth enough due to the inverse operator. So we smooth out the vector fields so that our final metric would be

$$G = K_\sigma * \begin{pmatrix} 1 + |\widehat{\partial_{x_1}I}| & 0 \\ 0 & 1 + |\widehat{\partial_{x_2}I}| \end{pmatrix}^{-1}.$$

```
Vx=1./(1+abs(Vx));
Vy=1./(1+abs(Vy));
Vx=gaussblur(Vx,9);
Vy=gaussblur(Vy,9);
```

Then the covariance matrix of the kernel is $HH' = 2tG$ and we perform iterated kernel smoothing with K_H . We performed 100 iterations with $t = 0.5$ which gives the right images in Figure 4. Notice

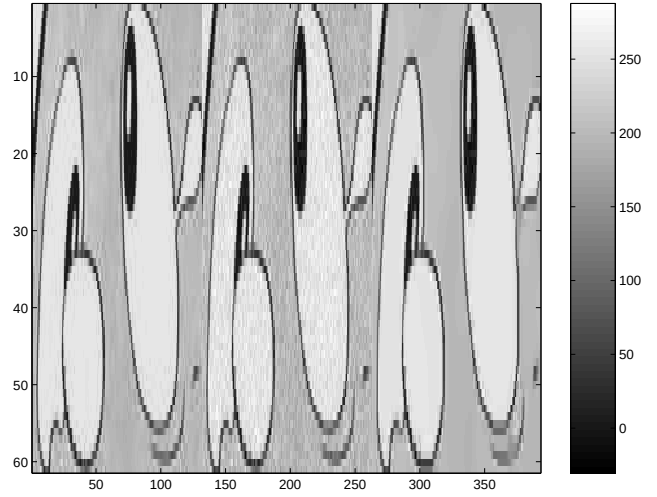


Figure 4: Left: original image, middle: noisy image, right: anisotropic smoothing of noisy image. See how it reduces noise dramatically without blurring the geometric features.

how it clears noise but preserves all the geometric features.

3. *M-estimation*. There is a slightly different approach to edge preserving smoothing. Given additive model

$$I(x) = \mu(x; \sigma) + \epsilon(x)$$

for image intensity, the *M*-estimator of unknown parameter σ is defined as a minimizer

$$\hat{\sigma} = \arg \min_{\sigma} \int \rho(I(x) - \mu(x; \sigma)) dx$$

In the case $\rho(z) = z^2$, we have the usual least squares estimation (LSE). The motivation for *M*-estimation is the (LSE) is not robust enough in the presence of outliers. The usual choice for the score function is that $\rho(z)/z^2 < \infty$ as $z \rightarrow \infty$.

For edge detecting problem, zero-one score function $\rho(z) = \mathbf{1}_{|z|>k}(z)$ has been used (Chen *et al.*, 2001, CVPR). Chu *et al.* (1998, JASA) used $\rho(z) = -\exp(-z^2/2\sigma^2)/\sigma$ and $\mu(x, \sigma) = \mu(x)$ which downweights image intensity far from the estimate. Huber in Robust statistics 1981 used $\rho(z) = z^2$ if $|z| \leq k$ and $2k|z| + k^2$ if $|z| > k$.

The main problem is formulating kernel smoothing in the context of the *M*-estimation problem which would give a statistical justification for this adaptive kernel smoothing. This requires figuring out the variational version of the kernel smoothing and diffusion smoothing. See Cai's Weighted Nadaraya-Watson Regression Estimation for reference and

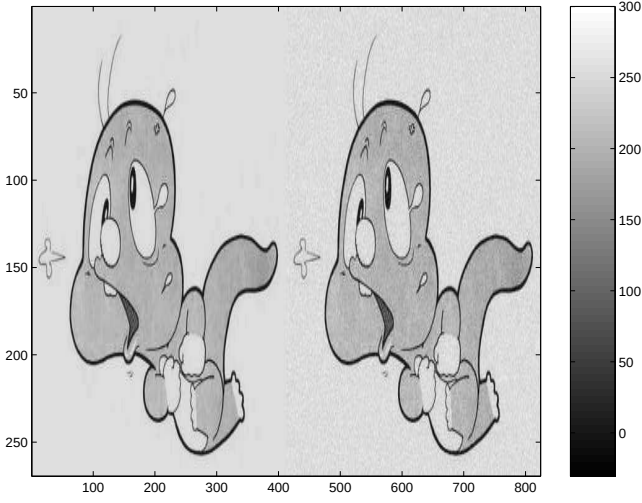


Figure 5: Left: original image, Right: noisy image

Fan's book. Consider the following discrete minimization problem

$$\hat{\beta}(x) = \arg \min_{\beta_j} \sum_{i=1}^n K_{\sigma}(x-x_i) \left[Y(x_i) - \sum_{j=0}^p \beta_j (x-x_i)^j \right]^2$$

It can be shown that the Nadaraya-Watson estimator $\hat{\mu}_{\sigma} = \widetilde{K}_{\sigma} * Y$ is the minimizer when $p = 0$. The continuous version would be obviously

$$K_{\sigma} * Y(x) = \arg \min_{\beta_0} \int K_{\sigma}(x-y) [Y(y) - \beta_0]^2 dy$$

This should be trivially true. Then the anisotropic version should be also correct.

Theorem.

$$K_H * Y(x) = \arg \min_{\beta} \int K_H(x-y) [Y(y) - \beta]^2 dy.$$

Then it follows that we can extend this result to manifolds so that we have the following most general theorem.

Theorem. For C^2 manifolds $\partial\Omega$ with the Laplace-Beltrami operator Δ with eigenvalue λ_j and eigenfunction ψ_j , the solution to diffusion equation

$$\partial_t f = \Delta f$$

with initial value $f(x, 0) = Y(x)$ is the minimizer

$$K_{\sqrt{2t}} * Y(x) = \arg \min_{\beta} \int_{\partial\Omega} K_{\sqrt{2t}}(x, y) [Y(y) - \beta]^2 d\mu(y).$$

where heat kernel is given by

$$K_{\sigma}(x, y) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} \psi_j(x) \psi_j(y).$$

Hence the diffusion and kernel smoothing on manifold can be viewed as the local polynomial regression. :)