Bayesian Approach to Anisotropic Diffusion

Jianzhong Wang
Sam Houston State University
Huntsville, Texas

International Conference
on Applicable Harmonic Analysis:
Approximation and Application
June 2006   Beijing, China

Mathematical Model

We assume the noise function is in the following model:

$$u_0(x) = v(x) + \xi(x), \quad x \in \Omega \subset \mathbb{R}^n$$

We introduce *maximum a posterior estimate* (MAP) method for reducing noise on functions and then develop the corresponding algorithms to realize the denoise processing.

Outline

- Introduction
- Posterior Energy in MAP Method
- Solution of MAP Estimate
- Directed Diffusion Kernel
- Content Dependent Filter Algorithm

Posterior Distribution of Discrete Data

The posterior distribution is obtained the modification of prior distribution using the likelihood function. Let S be the target data set.

- *Likelihood function* on the target set S is the conditional probability of u_0 given a target u:

$$u \longmapsto P(u_0|u)$$

- *prior distribution* $\Pi(u)$ is assigned on S to characterize the data features.
- Then *posterior distribution* of a target data u is

$$u \longmapsto \frac{\Pi(u)P(u_0|u)}{\sum_{z\in S}\Pi(z)P(u_0|z)}.$$

MAP estimate on Discrete data set

- The MAP method estimates the target v as the *mode* of the posterior:

$$v^* = \arg\max_{u \in S} \frac{\Pi(u) P(u_0|u)}{\sum_{z \in S} \Pi(z) P(u_0|z)}$$

- Note that the denominator above does not impact the maximization. Hence,

$$v^* = \arg\max_{u \in S} \Pi(u) P(u_0|u)$$

Posterior Energy Function

The *Gibbsian form* of the posterior $\Pi(u)P(u_0|u)$ yields the *posterior energy function*:

$$u \longmapsto H(u) + H(u, u_0),$$

where $H(u) = -\ln \Pi(u)$ is the *prior energy function*, and $H(u, u_0) = -\ln P(u_0|u)$ is the *likelihood energy function*. Then we have

$$v^* = \arg\min_{u \in S} \left(H(u) + H(u, u_0)\right)$$

Variation Method

In variational method, a total energy functional $E(u)$ is assigned on the target function set, and $E(u)$ is a sum of an *internal energy* functional

$$E_i(u) = \lambda \int_\Omega \rho(\|\nabla u\|)dx$$

and an *external energy* functional

$$E_e(u) = \lambda \int_\Omega \rho(\|\nabla u\|)dx + \frac{1}{2}\int_\Omega (u - u_0)^2 dx$$

Then

$$E(u) = E_i(u) + E_e(u)$$

Solution of Variation Method

An *estimate* of the target data v is

$$v^* = \arg\min \left(E_i(u) + E_e(u) \right)$$

Comparison with Variational Method

- Bayesian method utilizes the information about the probabilistic mechanism of the target function set. Hence, if the statistical information of the target function set is full, we can set a relatively precise prior for the target function set, while the selections of ρ in variational method seem quite empirical, lacking in selection laws.
- The rigid external energy in the variational method is replaced by a “weak ”condition given in the likelihood energy function $H(u_0|u)$, which gives more flexibility to choose an estimate for the target data.

Our Motivation

- Establish the relation between the Bayesian method and the variational method
- Take the benefits of well-developed PDE theory to develop denoising algorithms in spatial space (local filters)
- Develop selection laws for parameters in estimator

MAP for Continuous Data

- Observation space: $L^2(\Omega), \quad \Omega \subset \mathbb{R}^n$
- Target function set:

$$W_0^{k,p}\left(\overline{\Omega}\right) = \left\{ u \in W^{k,p}\left(\overline{\Omega}\right); \quad \frac{\partial u}{\partial \vec{n}} = 0, \text{ on } \partial\Omega \right\}.$$

Likelihood Energy Functional

The likelihood distribution $H(u_0|u)$ can be obtained in a simple way. Assume the noise $\xi(x)$ is i.i.d. random variable with the distribution density μ. Then the likelihood energy density is

$$H(u_0|u) = -\int_\Omega \ln \mu(u(x) - u_0(x))dx.$$

Assume the noise $\xi(x)$ is a white noise with the standard deviation σ. Then

$$H(u_0|u) = \frac{1}{2\sigma^2} \int_\Omega |u(x) - u_0(x)|^2 dx.$$

Prior Energy Functional

We assume the features of target functions are characterized by their gradient fields, and the prior distribution $P(u)$ is favor for smoother functions. For a target function u, we write $p_u(x) = ||\nabla u(x)||$. Assume the gradient field is governed by a distribution density $\pi(p)$:

$$P(p_u(x) \leq y) = \int_0^y \pi(p)dp,$$

Then the prior energy density is

$$H(u) = \int_\Omega \rho(||\nabla u(x)||)dx,$$

MAP Estimate Formulation

- In general, we have

$$u^* = \arg \min_{u \in W_0^{1,2}(\overline{\Omega})} \int_\Omega \lambda ||\nabla u(x)|| dx + \frac{1}{2\sigma^2} \int_\Omega |u(x) - u_0$$

- An important target function set S is that the magnitudes of gradients are governed by (one-sided) Laplacian distribution.

$$S = \left\{ u \in W_0^{1,2}\left(\overline{\Omega}\right); \quad P(||\nabla u|| \leq y) = \int_0^y \lambda e^{-\lambda s} ds\right.$$

where $\lambda > 0$ is a normalization parameter.

MAP with Laplacian Prior

$$u^* = \arg \min_{u \in W_0^{1,2}(\overline{\Omega})} \int_\Omega \lambda ||\nabla u(x)|| dx + \frac{1}{2\sigma^2} \int_\Omega |u(x) - u_0(x)|^2$$

A Basic Result from Variational Calculus

The Gâteaux derivative of a functional $E(u)$ in the direction v is defined by

$$dE(u;v) := \lim_{\lambda \to 0} \frac{E(u + \lambda v) - E(u)}{\lambda}.$$

The following result is well-known.
If $E(u)$ has a minimum at u^* and $dE(u^*;v)$ exists, then $dE(u^*;v) = 0$.

Solution of MAP Estimate

Let $c(p) = \frac{\rho'(p)}{p}$ and $g(s) = \ln \mu(s)$. Then the Euler-Lagrange equation of the MAP estimate is

$$\nabla \cdot (c(\|\nabla u\|)\nabla u) - g'(u - u_0) = 0.$$

The Euler-Lagrange equation of the special MAP estimator is

$$\lambda \nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|} \right) - \frac{1}{\sigma^2}(u - u_0) = 0.$$

Locally Differential Representation

Let $\mathbf{b}_1 = \begin{cases} (1,0), & \text{if } \nabla u = 0 \\ \frac{\nabla u}{||\nabla u||}, & \text{if } \nabla u \neq 0 \end{cases}$ be the gradient direction of u at x, and $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be an orthonormal basis if $\mathbb{R}^n$. Let $Lu = \nabla \cdot (c(\|\nabla u\|)\nabla u)$. Write $p = ||\nabla u||, \phi(p) = pc(p)$. Then

$$Lu = (\phi(p))' \frac{\partial^2 u}{\partial b_1^2} + c(p) \sum_{j=2}^{n} \frac{\partial^2 u}{\partial b_j^2}.$$

Steepest Descent Method

The steepest descent method to solve the equation above is described by the following evolution equation:

$$\frac{\partial u}{\partial t} = \nabla \cdot (c(\|\nabla u\|)\nabla u) - g'(u - u_0), \quad (x,t) \in \Omega \times \mathbb{R}^+,$$

$$\frac{\partial u}{\partial \vec{n}} = 0 \quad (x,t) \in \partial\Omega \times \mathbb{R}^+,$$

$$u(x,0) = u_0(x), \quad x \in \overline{\Omega}$$

Then the MAP estimate is the steady-state solution of the equation:

$$u^* = \lim_{t \to \infty} u(x,t).$$

Energy Inequality

Let $u(x,t)$ be the solution of the equation. Then $H(u(\cdot,t)|u_0) < H(u(\cdot,t')|u_0)$, if $t > t'$. That is, as t develops, the function $u(x,t)$ monotonously approximates the MAP estimate u^*.

Constrained Approach

Let

$$V = \left\{ u \in \overline{W}^{2,2}(\overline{\Omega}); \quad \int_\Omega |u(x) - u_0(x)|^2 dx = \sigma^2. \right\}$$

Then an estimate u^* of the target function v is

$$u^* = \arg\min_{u \in V} H(u).$$

The steepest descent method to find u^* is described by the homogeneous equation:

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(c(\|\nabla u) \|\nabla u \right)$$

Stop-time in Constrained Approach

Let $u(x,t)$ be the solution of the equation. Then $H(u(\cdot,t)) < H(u(\cdot,t')), if\ t > t'$. That is, as t develops, the function $u(x,t)$ monotonously approximates to a constant, that is, $\lim_{t\to\infty} u(x,t) = \text{const}$. Assume that the target function v is not a constant. Then there exists a stop-time t^* such that

$$\int_\Omega |u(x,t^*) - u_0(x)|^2 dx = \sigma^2.$$

Regularization of Directed Diffusion

We regularize the equation to

$$\frac{\partial u}{\partial t} = \frac{\lambda}{\sqrt{\epsilon + \|\nabla u\|^2}} \Delta_g u, \qquad (x,t) \in \Omega \times \mathbb{R}^+, \quad (1)$$

$$\frac{\partial u(x,t)}{\partial \vec{n}} = 0, \qquad (x,t) \in \Omega \times \mathbb{R}^+,$$

$$u(x,0) = u_0(x), \qquad x \in \Omega,$$

where $\epsilon > 0$ is a small number.

Time Discretization

$$u^{n+1} = u^n + \frac{\lambda s}{\sqrt{\epsilon + \|\nabla u^n\|^2}} \Delta_g u^n + \bigcirc(s)$$

We want to find a kernel $K^n(x, y)$ such that

$$u^{n+1}(x) = \int_\Omega K^n(x, y) u^n(y) dy.$$

Gaussian Kernel

It is well-known that the Laplacian $\Delta = \sum_{j=1}^{n} \frac{\partial^2}{\partial x_j^2}$ is the infinitesimal of the semi-group of G^t, where G^t is represented by the Gaussian kernel.

$$G^t u(x) = \frac{1}{\left(\sqrt{2\pi t}\right)^n} \int_{\Omega_t} \exp\left(-\frac{\|x-y\|^2}{2t}\right) u(y)dy + \circ(t^n).$$

Directed Diffusion Kernel

For $s > 0, \alpha > 0$, and $g \in W^{k,p}(\overline{\Omega})$, we define the directed diffusion kernel by

$$F^{(\alpha)}_{g,s}(x,y) = \exp\left(-\frac{||x-y||^2}{s} - \frac{|< \nabla g(x), x-y >|^2}{s^\alpha} \right)$$

It is clear that, when $||x-y||$ is fixed, the larger is $\|\nabla g(x)\|$, the smaller is $F_{g,s}(x,y)$. Hence, the directed diffusion is favor for points at the same level set.

Normalization of Directed Diffusion Kernel (1)

Let $p_s(x) = \int_{\Omega_s} F_{g,s}^{(\alpha)}(x,y)dy$. Then

$$p_s(x) = \sqrt{(\pi s)^n} q_s(x) \left(1 + \circ(s^m)\right), \quad s \longrightarrow 0, \quad \forall x \in \overline{\Omega}$$

where $q_s(x) = \sqrt{\frac{s^{\alpha-1}}{s^{\alpha-1} + \|\nabla g(x)\|^2}}$ and $m > 0$ be an arbitrary large integer.

Normalization of Directed Diffusion Kernel (2)

We define the normalization of $F_{g,s}$ called $A_{g,s}^{(\alpha)}$ by

$$A_{g,s}^{(\alpha)}(f)(x) = \frac{1}{p_s(x)} \int_{\Omega_s} F_{g,s}^{(\alpha)}(x,y) f(y) dy, \quad x \in \Omega.$$

Then $A_{g,s}^{(\alpha)}$ is strongly convergent to the identity I:

$$\lim_{s \longrightarrow 0} ||A_{g,s}^{(\alpha)} f - f||_{L^2(\Omega)} = 0.$$

Infinitesimal of Directed Diffusion Kernel

Let $L_{g,s}^{(\alpha)} = \frac{A_{g,s}^{(\alpha)} - I}{s}$. Then for $\alpha > 1$ we have

$$A_{g,s}^{(\alpha)} f(x) = f(x) + \frac{s}{4} \Delta_g f(x) + \bigcirc \left(s^{\max(\alpha, 3/2)} \right)$$

and therefore

$$\lim_{s \to 0} L_{g,s}^{(\alpha)} f = \frac{1}{4} \Delta_g f, \qquad f \in W_0^{2,2}(\overline{\Omega}).$$

Local Diffusion Kernel

$$A_s^n(x,y) = \frac{1}{\sqrt{\pi^n s^{n+1}}} \exp\left(-\frac{\|x-y\|^2}{s} - \frac{| < \nabla u^n, x-y > |^2}{s^2} \right)$$

$$A_s^n f(x) = \int_{\Omega_s} A_s^n(x,y) f(y) dy, \quad x \in \overline{\Omega}.$$

Solution Represented by Kernel

Let u^n be the solution of the time-discrete equation with $\epsilon = s$. Then

$$u^{n+1} = \left(\left(1 - \frac{4\lambda s}{\sqrt{s + \|\nabla u^n(x)\|^2}} \right) I + 4\lambda A_s^n \right) u^n + \bigcirc\,(s$$

A similar kernel can be derived for the inhomogeneous equation.

Spatial discretization: Filtering

We denote $x^i = \frac{1}{N}(i_{1,} \cdots, i_n)^T$. The discrete function u defined on the grid set $\left\{x^i; \quad i \in \mathbb{Z}^n\right\}$ is denoted by $\vec{u} = u(x)$. A k-window around x^i is the set

$$Q_i = \{x^j; \quad -k \leq x^j - x^i \leq k\}.$$

We write $Q = Q_0$ and then $Q_i = x^i + Q$. We said a filter F has support Q, if

$$F\vec{u}(x^i) = \sum_{j \in Q_i} w_i^j \vec{u}(x^j).$$

Discrete Gradients

Let $\vec{n}$ denote the direction of $y - x$. Then

$$|< \nabla u, y - x >|^2 = \left|\frac{\partial u(x)}{\partial \vec{n}}\right|^2 \|y - x\|^2 .$$

In the discrete form, we use

$$|f(y) - f(x)|^2 \approx |< \nabla u, y - x >|^2$$

Approximation of Directed Diffusion Kernel

$$A^{(2)}_{u,s}\left(\vec{u}\right)\left(x^i\right) =$$

$$\frac{1}{c(i)}\sum_{j\in Q_i}\exp\left(-\frac{||x^j-x^i||^2}{s}-\frac{|u(x^j)-u(x^i)|^2}{s^2}\right)\vec{u}(x^j).$$

where

$$c(i)=\sum_{j\in Q_i}\exp\left(-\frac{||x^j-x^i||^2}{s}-\frac{|u(x^j)-u(x^i)|^2}{s^2}\right).$$

The processing is an inhomogeneous Markov processing.

Estimate of Noise STD

$$\sigma \simeq \operatorname*{med}_{x^i \in \Omega} \left(\frac{1}{\sqrt{2n}} \sqrt{\sum_{|i-j|=1} \left| \vec{u}_0(x^j) - \vec{u}_0\left(x^i\right) \right|^2} \right) / 0.9549$$

Comparing to [Donoho and Johnstone]:

$$\sigma \simeq \operatorname*{med}_{m} \left(w_m^J \right) / 0.6745.$$

Estimate of Normalization Parameter λ

Let

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-r^2} dr.$$

The gradient magnitude of u_0 is

$$f_{u_0}(y) = \frac{\lambda}{2} e^{\lambda(|y|+\lambda\sigma^2)} \left(1 - \mathrm{erf}\left(\frac{|y| + 2\lambda\sigma^2}{2\sigma}\right)\right) + \frac{\lambda}{2} e^{-\lambda(|y|-\lambda\sigma^2)} \left(1 - \tau \mathrm{erf}\left(\frac{-|y| + 2\lambda\sigma^2}{2\sigma}\right)\right).$$

where $\tau = \mathrm{sign}\left(-|y| + 2\lambda\sigma^2\right)$. We then can estimate λ by the histogram of ∇u_0

END

THANK YOU