Lecture 7:

Total Variation Denoising

Calculus of Variations

- A functional is a mathematical operation that maps a function to a real number.
- One example of a functional is a definite integral:

$$\int_a^b g(x)dx$$

- The calculus of variations is the study of finding the maxima & minima of functionals.
- Many applications:
  - Path planning: minimize the distance between 2 points
  - Physics: systems seek out lowest energy state
  - Image processing: minimize noise & blur

Double Integral Shorthand

- In image processing, many of our problems will involve minimizing the double integral of an image u:

$$\iint_{\Omega} g\big(u(x,y)\big)\, dx\, dy$$

where Ω is our rectangular domain.

- To save a little bit of writing, we will vectorize our variables and just write a single integral sign:

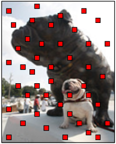
$$\boxed{\int_{\Omega} g(u(\vec{x}))\, d\vec{x}}$$

Variational Methods

- All things in nature seek out a lower energy state.
- A variational method (also known as energy or PDE-based methods) prescribes an energy to an image.

High energy = "Bad" image | *Low energy = "Good" image*

Bad image f



Energy Minimization

Good image u

- By making the energy smaller, we improve the image.
- We want an energy that describes the "quality" of an image.

First Variation of Energy

- For a functional of the form

$$E[u] = \int_{\Omega} g(u, u_x, u_y) d\vec{x}$$

- According to *(Lagrange, 1755)*, the first variation of energy is given by

$$\nabla E = \frac{\partial g}{\partial u} - \frac{\partial}{\partial x}\frac{\partial g}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial g}{\partial u_y}$$

- Note we treat u and its derivatives as separate variables in the calculation.
- ∇E acts like a gradient:
  - Extrema occur where $\nabla E = 0$ *(Euler-Lagrange Equation)*
  - ∇E points in the direction of maximum positive rate of change

Steepest Descent

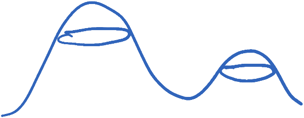
- One way to find the minimum of $E[u]$ is to take small step of size Δt in the direction of $-\nabla E$.

$$\boxed{u^{n+1} = u^n + (\Delta t)(-\nabla E)}$$

- Introducing an artificial time parameter, we can rewrite this as a PDE.

$$\frac{u^{n+1} - u^n}{\Delta t} = -\nabla E$$

$$\frac{\partial u}{\partial t} = -\nabla E$$



- Will this algorithm always guarantee that we find the global minimum?

Convexity

- We say our energy E[u] is convex if for any 2 function u and v it satisfies:

$$\boxed{E[tu + (1 - t)v] \leq tE[u] + (1 - t)E[v]}$$

- Fact: The local minimum of a convex function is also a global minimum.

- So we want to choose convex energies, or else we might get stuck at a local min.

H1 Regularization

- A noise point is a sudden jump, so we expect $||\nabla u||$ to be large near noise.
- A classical regularization to smooth a surface is the H^1 norm.

$$\boxed{E[u] = \int_{\Omega} ||\nabla u||^2 d\vec{x}}$$

where Ω is the image domain.

First Variation of Energy

- Calculate the first variation of $E[u] = \frac{1}{2}\int_{\Omega} ||\nabla u||^2 dx$

$$g(u, u_x, u_y) = \frac{1}{2}\|\nabla u\|^2 = \frac{1}{2}\left(\sqrt{u_x^2 + u_y^2}\right)^2 = \frac{1}{2}(u_x^2 + u_y^2)$$

$$\nabla E = \frac{\partial g}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial g}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial g}{\partial u_y}\right)$$

$$= \frac{1}{2}\left(0 - \frac{\partial}{\partial x}(2u_x) - \frac{\partial}{\partial y}(2u_y)\right)$$

$$= \frac{1}{2}(-2u_{xx} - 2u_{yy}) = -\frac{2}{2}\Delta u = \boxed{-\Delta u}$$

$\Delta u = u_{xx} + u_{yy}$

Laplacian.

Steepest Descent

- So for our energy, the steepest descent gives

$$\boxed{\frac{\partial u}{\partial t} = \Delta u}$$

- This is the heat equation!

Minimize the H1 norm

$$\int_{\Omega} |\nabla u|^2 d\vec{x}$$

Evolve the heat equation

$$\frac{\partial u}{\partial t} = \Delta u$$

Boundary Conditions

- Assume Neumann boundary conditions $\frac{\partial u}{\partial \vec{n}} = 0$.

The Variational Approach

Create energy $E[u]$

Calculate first variation ∇E

Minimize by steepest descent $\frac{\partial u}{\partial t} = -\nabla E$

Discretize and code
$$u^{n+1} = u^n + \Delta t(-\nabla E)$$

- The challenge is to cook up better energies and faster minimization schemes.

Data Fidelity

- The heat equation blurs the image. If left to run for a long time, it will eventually give u=const.
- To prevent this, we can add a data fidelity term to our energy to keep our image close to the original image f.

$$\min E[u|f] = \underbrace{\int_\Omega ||\nabla u||^2 d\vec{x}}_{\text{Regularization Term}} + \lambda \underbrace{\int_\Omega (u-f)^2 d\vec{x}}_{\text{Data Fidelity Term}}$$

- The parameter λ balances the two terms. We call λ the fidelity weight or a Lagrange multiplier.

Data Fidelity

$$\min E[u|f] = \int_{\Omega} ||\nabla u||^2 d\vec{x} + \lambda \int_{\Omega} (u - f)^2 d\vec{x}$$

- The first variation is easy to modify.

$$\boxed{\nabla E = -2\Delta u + 2\lambda(u - f)}$$

- So the PDE becomes

$$\boxed{\frac{\partial u}{\partial t} = \Delta u - \lambda(u - f)}$$

- Even with the fidelity term, the H^1 model blurs the edges.

Regularization

- Consider regularization of the form $\int_{\Omega} ||\nabla u||^p d\vec{x}$.
- How would this regularization perform?

Total Variation

- The choice p=1 preserves both sharp and fuzzy edges while removing noise.
- *(Rudin-Osher-Fatemi, 1992)* proposed the Total Variation (TV) denoising model.

$$\min E_{TV}[u|f] = \int_{\Omega} ||\nabla u|| d\vec{x} + \lambda \int_{\Omega} (u-f)^2 d\vec{x}$$

- TV is the most famous and widely used image restoration model. Its simplicity and adaptability makes it very useful.

Total Variation Energy

- Calculate the first variation of the TV energy

$$E_{TV}[u|f] = \int_{\Omega} ||\nabla u|| d\vec{x} + \lambda \int_{\Omega} (u-f)^2 d\vec{x}$$

Total Variation

$$\min E_{TV}[u|f] = \int_{\Omega} |\nabla u| d\vec{x} + \lambda \int_{\Omega} (u-f)^2 d\vec{x}$$

- The first variation of energy is

$$\boxed{\nabla E = -\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) + 2\lambda(u-f)}$$

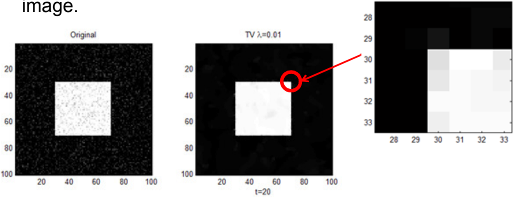
$$= -\frac{u_{xx}{u_y}^2 - 2u_x u_y u_{xy} + u_{yy}{u_x}^2}{\left({u_x}^2 + {u_y}^2\right)^{3/2}} + 2\lambda(u-f)$$

- The energy is convex, so we can use steepest descent to evolve the PDE

$$\frac{\partial u}{\partial t} = \frac{u_{xx}{u_y}^2 - 2u_x u_y u_{xy} + u_{yy}{u_x}^2}{\left({u_x}^2 + {u_y}^2\right)^{3/2}} - 2\lambda(u-f)$$

Total Variation Minimization

- TV minimization is good for removing noise from an image.

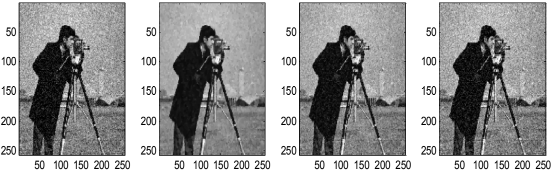


- TV minimization tends to round off the corners of objects.

Parameter Selection

- $\min E_{TV}[u|f] = \int_{\Omega} ||\nabla u|| d\vec{x} + \lambda \int_{\Omega} (u-f)^2 d\vec{x}$
- The choice of λ is important.
  - λ too small -> u will be blurry
  - λ too large -> u will resemble f
- Generally, the more noise we have the smaller we set λ.
- There are automatic methods for parameter selection such as the L-curve method *(Vogel, 1996)*.
- Some people recommend using a spatially varying λ *(Gilboa-Sochen-Zeevi, 2006)*.

Noisy λ = 0.01 λ = 0.1 λ = 1



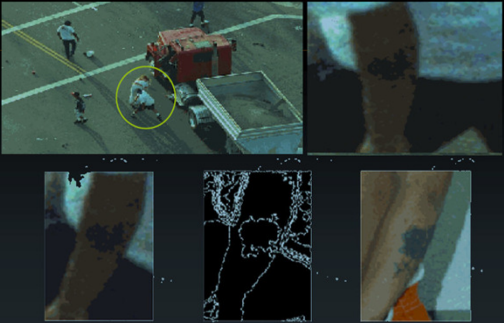
Parameter Selection

- As λ gets larger, the result resembles the original.

TV-regularized denoising with increasing values of λ.

(Getreuer, 2012)

The LA riot in 1991 and the rose tattoo



Theory

- Existence: A unique minimizer exists in the space of Bounded Variation (BV), which is where the TV norm is finite.
- Maximum Principle: The values of the minimizer u will not go beyond the max/min of the original image f.
- Co-Area Formula: The TV norm is equal to the perimeter of the level sets $\int |\nabla u| d\vec{x} = \int_{-\infty}^{\infty} Per\{u = \beta\} d\beta$ *(Giusti, 1984)*.

Minimization

- The technique we used to minimize TV is called Explicit Forward Euler.
- There exist much faster methods for minimizing the TV energy.

FASTER

  - Implicit solvers *(Chambolle-Levine-Lucier, 2011)*
  - Newton methods *(Ng-Qi-Yang-Huang, 2007)*
  - TV filter *(Chan-Osher-Shen, 2001)*
  - Duality *(Chambolle, 2005)*
  - Graph cuts *(Darbon-Sigelle, 2005)*
  - Bregman iteration *(Goldstein-Osher, 2009)*