Lab 1

Total Variation and Image Processing

Lab Objective: *Minimizing an energy functional is equivalent to solving the resulting Euler-Lagrange equations. We introduce the method of steepest descent to solve these equations, and apply this technique to a denoising problem in image processing.*

The Gradient Descent method

Consider an energy functional $E[u]$, defined over a collection of admissible functions $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. We suppose the energy functional E has the form

$$E[u] = \int_\Omega L(x, u, \nabla u)\, dx$$

where $L = L(x, y, w)$ is a function $\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$. A standard result from the calculus of variations states that a minimizing function u^* satisfies the Euler-Lagrange equation

$$L_y - \operatorname{div}(L_w) = 0. \tag{1.1}$$

This equation is typically an elliptic PDE, possessing boundary conditions associated with restrictions on the class of admissible functions u. To more easily compute (1.1), we may consider a related parabolic PDE:

$$\begin{aligned} u_t &= -(L_y - \operatorname{div} L_w), \\ u(x, t = 0) &= u_0(x). \end{aligned} \tag{1.2}$$

A steady state solution of (1.2) does not depend on time, and thus is a solution of the Euler-Lagrange equation. In practice, it is often easier to evolve an initial guess u_0 using (1.2), stopping whenever its steady state is well-approximated, then to solve (1.1) directly.

Example 1.1. Consider the energy functional

$$E[u] = \int_\Omega \|\nabla u\|^2\, dx,$$

where the class of admissible functions u satisfy appropriate Dirichlet conditions on $\partial\Omega$. The minimizing function u^* satisfies the Euler-Lagrange equation

$$-\operatorname{div}\nabla u = -\triangle u = 0.$$

The related PDE (called the gradient descent flow) is the well-known equation

$$u_t = \triangle u$$

describing heat flow.

The previous example brings to mind an interesting question: The Euler-Lagrange equation could equivalently be described as $\triangle u = 0$, leading to the PDE $u_t = -\triangle u$. Since this (backward) heat equation is ill-posed, it will not be helpful in our search for a steady-state.

Let us take the time to put (1.2) on a more rigorous footing. Recalling the derivation of the Euler-Lagrange equation, we note that

$$\begin{aligned}\delta E(u;v) &= \frac{d}{dt}E(u+tv)\Big|_{t=0}, \\ &= \int_\Omega (L_y(u) - \operatorname{div} L_w(u))v\,dx\end{aligned}$$

for each u and each admissible perturbation v. We can now employ the Cauchy-Schwarz inequality:

$$\begin{aligned}|\delta E(u;v)| &= |\langle L_y(u) - \operatorname{div} L_w(u), v\rangle_{L^2(\Omega)}|, \\ &\leq \|L_y(u) - \operatorname{div} L_w(u)\| \cdot \|v\|,\end{aligned}$$

with equality iff $v = \alpha u$ for some $\alpha \in \mathbb{R}$. This implies that the "direction" $v = L_y(u) - \operatorname{div} L_w(u)$ maximizes $\delta E(u)$. Similarly,

$$v = -(L_y(u) - \operatorname{div} L_w(u))$$

points in the direction of steepest descent, and the flow described by (1.2) tends to move toward a state of lesser energy.

Example: Finding the curve that surface of revolution minimizes surface area

Consider the collection of smooth curves defined on $[a, b]$, with fixed end points $y(a) = y_a$, $y(b) = y_b$. The surface obtained by revolving a curve $y(x)$ about the x-axis has area given by the functional

$$A[y] = \int_a^b 2\pi y\sqrt{1 + (y')^2}\,dx.$$

The Euler-Lagrange equation is

$$\begin{aligned}0 &= 1 - y\frac{y''}{1 + (y')^2}, \\ &= 1 + (y')^2 - yy'',\end{aligned} \tag{1.3}$$

and the resulting gradient descent flow is given by

$$\begin{aligned} u_t &= -1 - (y')^2 + yy'', \\ u(a,t) &= y_a, \quad u(b,t) = y_b, \\ u(x,0) &= g(x), \end{aligned} \tag{1.4}$$

where $g(x)$ is an appropriate initial guess.

Consider a second-order order discretization in space, with a simple forward Euler step in time. Let us impose the conditions $y(-1) = 1$, $y(1) = 7$. We begin by creating a grid to approximate the solution on:

```
import numpy as np

a, b = -1, 1.
alpha, beta = 1., 7.
####  Define variables x_steps, final_T, time_steps  ####
delta_t, delta_x = final_T/time_steps, (b-a)/x_steps
x0 = np.linspace(a,b,x_steps+1)
```

Often there is a stability condition required for a numerical scheme. One that is common for this discretization is that $\frac{\triangle t}{(\triangle x)^2} \leq \frac{1}{2}$. We continue by checking that this condition is satisfied, and setting our initial guess to be the straight line connecting the end points.

```
# Check a stability condition for this numerical method
if delta_t/delta_x**2. > .5:
    print "stability condition fails"

u = np.empty((2,x_steps+1))
u[0]  = (beta - alpha)/(b-a)*(x0-a)  + alpha
u[1] = (beta - alpha)/(b-a)*(x0-a)  + alpha
```

Finally, we define the right hand side of our difference scheme, and time step until we achieve a desired accuracy.

```
def rhs(y):
    # Approximate first and second derivatives to second order accuracy.
    yp = (np.roll(y,-1) - np.roll(y,1))/(2.*delta_x)
    ypp = (np.roll(y,-1) - 2.*y + np.roll(y,1))/delta_x**2.
    # Find approximation for the next time step, using a first order Euler step
    y[1:-1] -= delta_t*(1. + yp[1:-1]**2. - 1.*y[1:-1]*ypp[1:-1])


# Time step until successive iterations are close
iteration = 0
while iteration < time_steps:
    rhs(u[1])
    if norm(np.abs((u[0] - u[1]))) < 1e-5: break
    u[0] = u[1]
    iteration+=1

print "Difference in iterations is ", norm(np.abs((u[0] - u[1])))
print "Final time = ", iteration*delta_t
```

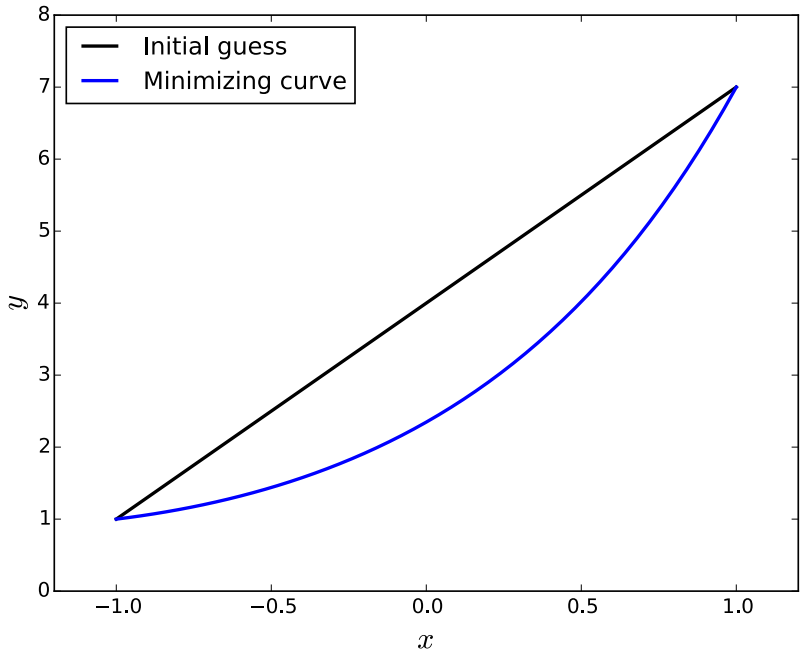


Figure 1.1: The solution of (1.3), found using the gradient descent flow (1.4).

Image Processing: A First Attempt

We represent a (greyscale) image by a function $u : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^2$. We use the following code to read an image into an array of floating point numbers, add some noise, and save the noisy image:

```
from numpy.random import random_integers, uniform, randn
import matplotlib.pyplot as plt
from matplotlib import cm
from scipy.misc import imread, imsave

imagename = 'baloons_resized_bw.jpg'
changed_pixels=40000
# Read the image file imagename into an array of numbers, IM
# Multiply by 1. / 255 to change the values so that they are floating point
# numbers ranging from 0 to 1.
IM = imread(imagename, flatten=True) * (1. / 255)
IM_x, IM_y = IM.shape

for lost in xrange(changed_pixels):
    x_,y_ = random_integers(1,IM_x-2), random_integers(1,IM_y-2)
    val =  .1*randn() + .5
    IM[x_,y_] = max( min(val,1.), 0.)
imsave(name=("noised_"+imagename),arr=IM)
```

We note that a color image can be represented by three functions u_1, u_2, and u_3. In this lab we will work with black and white images, but the techniques can easily

be used on more general images.

Here is our first attempt at denoising: given a noisy image f, we look to find a denoised image u minimizing the energy functional

$$E[u] = \int_\Omega L(x, \nabla u, u)\, dx, \tag{1.5}$$

where

$$\begin{aligned} L(x, \nabla u, u) &= \frac{1}{2}(u-f)^2 + \frac{\lambda}{2}|\nabla u|^2, \\ &= \frac{1}{2}(u-f)^2 + \frac{\lambda}{2}(u_x^2 + u_y^2)^2. \end{aligned}$$

This energy functional penalizes 1) images that are too different from the original noisy image, and 2) images that have a large derivatives. The minimizing denoised image u will balance these two different costs.

Solving for the original denoised image u is a difficult inverse problem; some information about the image is irretrievably lost when the noise is introduced. A priori information can however be used to guess at the structure of the original image. For example, in this problem λ represents our best guess on how much noise was added to the image. λ is known as a regularization parameter in inverse problem theory.

The Euler-Lagrange equation determined from (1.5) is

$$\begin{aligned} L_u - \text{div } L_{\nabla u} &= (u-f) - \lambda \triangle u, \\ &= 0. \end{aligned}$$

The resulting gradient descent flow is then given by

$$\begin{aligned} u_t &= -(u - f - \lambda \triangle u), \\ u(x,0) &= f(x). \end{aligned} \tag{1.6}$$

Let u_{ij}^n represent our approximation to $u(x_i, y_j)$ at time t_n. We will approximate u_t with a forward Euler difference, and $\triangle u$ with centered differences:

$$\begin{aligned} u_t &\approx \frac{u_{ij}^{n+1} - u_{ij}^n}{\triangle t}, \\ u_{xx} &\approx \frac{u_{i+1,j}^n - 2u_{ij}^n + u_{i-1,j}^n}{\triangle x^2}, \\ u_{yy} &\approx \frac{u_{i,j+1}^n - 2u_{ij}^n + u_{i,j-1}^n}{\triangle y^2}. \end{aligned}$$

Problem 1. Using $\triangle t = 1e-3, \lambda = 40, \triangle x = 1$, and $\triangle y = 1$, implement the numerical scheme mentioned above to obtain a solution u. (So $\Omega = [0, n_x] \times [0, n_y]$, where n_x and n_y represent the number of pixels in the x and y dimensions, respectively.) Take 250 steps in time. Compare your results with Figure 1.3.

Hint: use the function `np.roll` to compute the spatial derivatives. Consider the following:

Original image

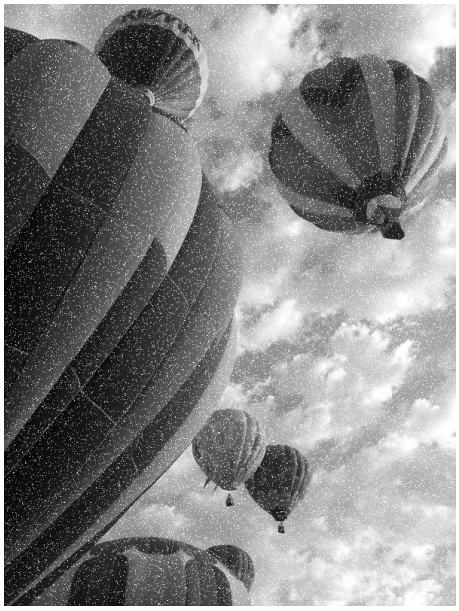


Image with white noise

Figure 1.2: Noise.

```
u_xx = np.roll(u,-1,axis=0) - 2*u + np.roll(z,1,axis=0)
```

Image Processing: Total Variation Method

We represent an image by a function $u : [0,1] \times [0,1] \to \mathbb{R}$. A C^1 function $u : \Omega \to \mathbb{R}$ has bounded total variation on Ω ($BV(\Omega)$) if $\int_\Omega |\nabla u| < \infty$; u is said to have total variation $\int_\Omega |\nabla u|$. Intuitively, the total variation of an image u increases when noise is added.

The total variation approach was originally introduced by Ruding, Osher, and Fatemi[1]. It was formulated as follows: given a noisy image f, we look to find a denoised image u minimizing

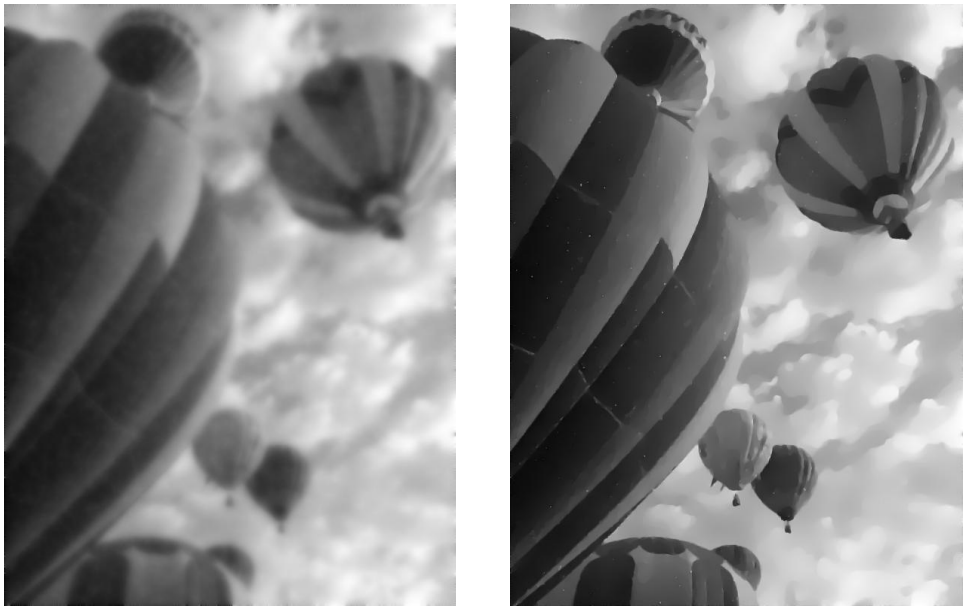
$$\int_\Omega |\nabla u(x)|\, dx \tag{1.7}$$

subject to the constraints

$$\int_\Omega u(x)\, dx = \int_\Omega f(x)\, dx, \tag{1.8}$$

$$\int_\Omega |u(x) - f(x)|^2\, dx = \sigma|\Omega|. \tag{1.9}$$

[1] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms", *Physica D.*, 1992.



Initial diffusion-based approach Total variation based approach

Figure 1.3: The solutions of (1.6) and (1.11), found using a first order Euler step in time and centered differences in space.

Intuitively, (1.7) penalizes fast variations in f - this functional together with the constraint (1.8) has a constant minimum of $u = \frac{1}{|\Omega|}\int_\Omega u(x)\,dx$. This is obviously not what we want, so we add a constraint (1.9) specifying how far $u(x)$ is required to differ from the noisy image f. More precisely, (1.8) specifies that the noise in the image has zero mean, and (1.9) requires that a variable σ be chosen a priori to represent the standard deviation of the noise.

Chambolle and Lions proved that the model introduced by Rudin, Osher, and Fatemi can be formulated equivalently as

$$F[u] = \min_{u\in BV(\Omega)} \int_\Omega |\nabla u| + \frac{\lambda}{2}(u-f)^2\,dx, \tag{1.10}$$

where $\lambda > 0$ is a fixed regularization parameter[2]. Notice how this functional differs from (1.5): $\int_\Omega |\nabla u|$ instead of $\int_\Omega |\nabla u|^2$. This turns out to cause a huge difference in the result. Mathematically, there is a nice way to extend F and the class of functions with bounded total variation to functions that are discontinuous across hyperplanes. The term $\int |\nabla|$ tends to preserve edges/boundaries of objects in an image.

[2]A. Chambelle and P.-L. Lions, "Image recovery via total variation minimization and related problems", *Numer. Math.*, 1997.

The gradient descent flow is given by

$$\begin{aligned} u_t &= -\lambda(u-f) + \frac{u_{xx}u_y^2 + u_{yy}u_x^2 - 2u_x u_y u_{xy}}{(u_x^2+u_y^2)^{3/2}}, \\ u(x,0) &= f(x). \end{aligned} \tag{1.11}$$

Notice the singularity that occurs in the flow when $|\nabla u| = 0$. Numerically we will replace $|\nabla u|^3$ in the denominator with $(\epsilon + |\nabla u|^2)^{3/2}$, to remove the singularity.

Problem 2. Using $\triangle t = 1e-3, \lambda = 1, \triangle x = 1$, and $\triangle y = 1$, implement the numerical scheme mentioned above to obtain a solution u. Take 200 steps in time. Compare your results with Figure 1.3. How small should ϵ be?

Hint: To compute the spatial derivatives, consider the following:

```
u_x = (np.roll(u,-1,axis=0) -  np.roll(z,1,axis=0))/2
u_xx = np.roll(u,-1,axis=0) - 2*u + np.roll(u,1,axis=0)
u_xy = (np.roll(u_x,-1,axis=1) - np.roll(u_x,1,axis=1))/2.
```