Image Processing and Computer Graphics

Image Processing

Class 5

Variational Methods

Discrete vs. continuous energies

- So far we have only considered discrete energy minimization problems

  $u^* = \text{argmin}_u E(u)$

  with a <u>vector</u> $u \in \mathbb{R}^N$

- Alternative approach: continuous formulation

  $u^*(\mathbf{x}) = \text{argmin}_{u(\mathbf{x})} E(u(\mathbf{x}))$

  with a <u>function</u> $u(\mathbf{x}) : \Omega \to \mathbb{R}$

- In this case, the energy function becomes an **energy functional**
- Optimization is based on the **calculus of variation** (Variationsrechnung)
- Yields necessary condition(s) for a minimum: **Euler-Lagrange equations**
- Discretization postponed to these equations

Discrete or continuous energies?

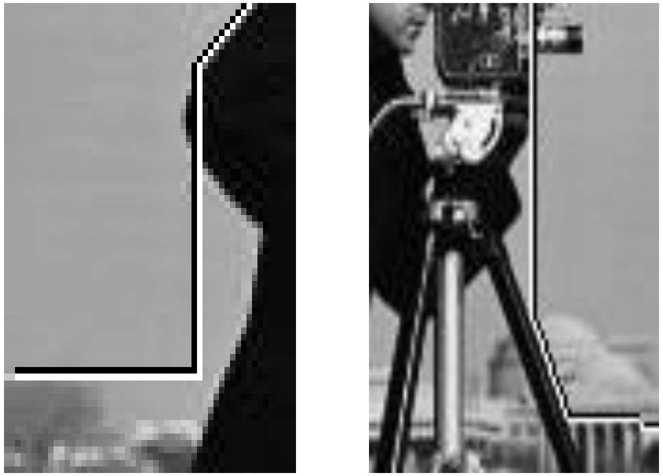
Advantage of continuous energies

- It is not ensured that a discrete energy is **consistent** with the continuous one. Typical example: measuring line lengths.

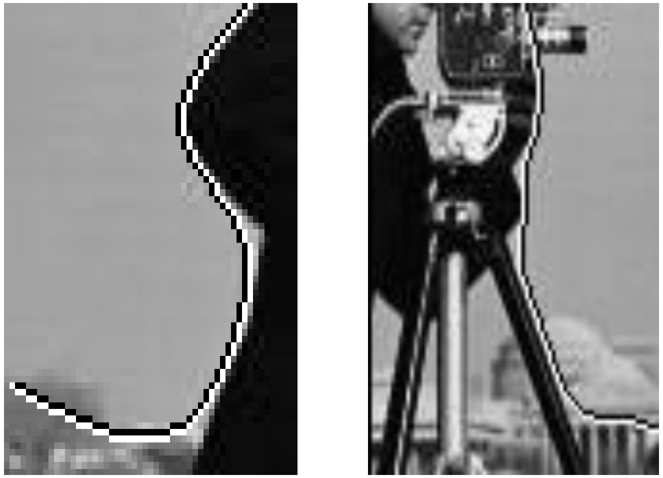
Disadvantage

- Gradients in direction of a function have to be computed.

However, this is not as difficult as it seems…



Optimum contours for a discrete energy



Optimum contours for a continous energy

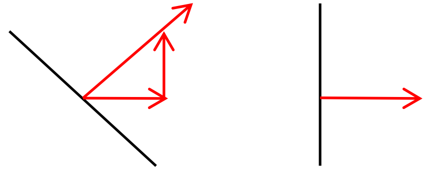
Author: Maria Klodt

Consistency

- Discretization of continuous quantities can lead to artifacts if done wrong
- For instance, we want the solution of a problem to be independent of the rotation of the grid (rotation invariance)
- A discretization is called **consistent** if it converges to the continuous model with finer grid sizes
- Example: discrete approximation of the gradient magnitude
  - Inconsistent discretization

    $$|\nabla u| \approx |u_{i+1,j} - u_{i-1,j}| + |u_{i,j+1} - u_{i,j-1}|$$

    There is an error independent of the grid size

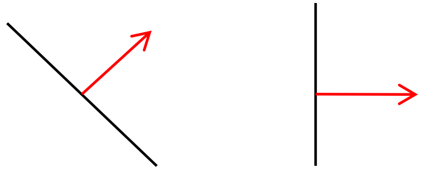


different magnitudes

  - Consistent discretization

    $$|\nabla u| \approx \sqrt{(u_{i+1,j} - u_{i-1,j})^2 + (u_{i,j+1} - u_{i,j-1})^2}$$

    All errors depend on the grid size



same magnitudes

Calculus of variation

- In the discrete setting we obtained a necessary condition for a minimum by computing the gradient of the function $E(u)$ with respect to the unknown vector $u \in \mathbb{R}^N$.

- This gradient $\frac{\partial E(u)}{\partial u}$ was computed by the tool of partial derivatives $\frac{\partial E(u)}{\partial u_i}$

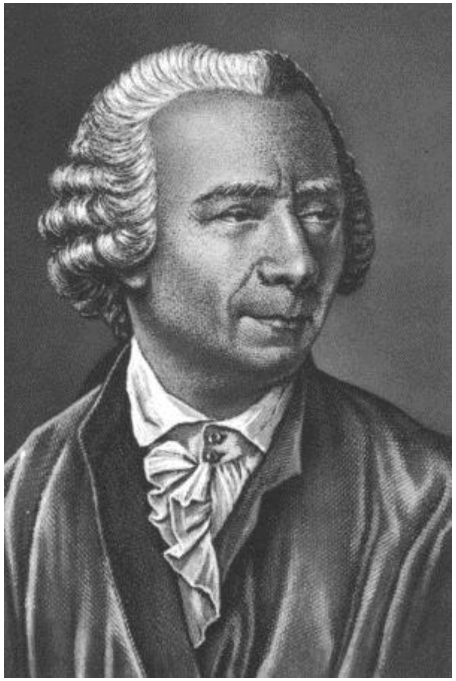
- How do we get the gradient of a functional $\frac{\partial E(u(\mathbf{x}))}{\partial u(\mathbf{x})}$ with respect to a function $u(\mathbf{x})$?

- The answer is the calculus of variation and the **Gâteaux derivative**.

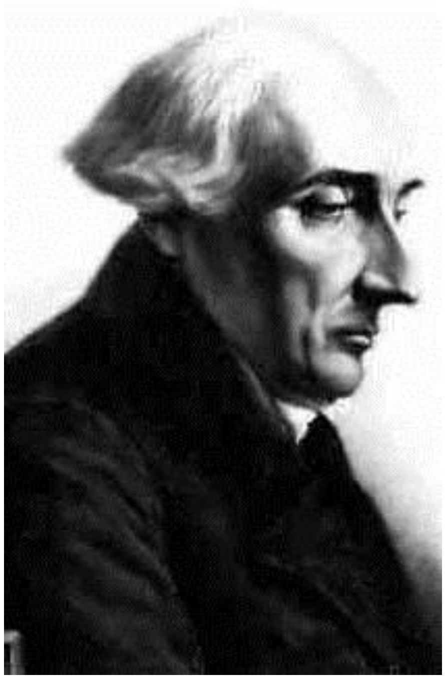
- It has been invented by the French mathematician René Gâteaux (who died in World War I) and has been published posthumously 1919.

- Similar ideas date back to the works of Leonhard Euler and Joseph-Louis Lagrange in the 18th century. The necessary conditions are hence called Euler-Lagrange equations.

Euler and Lagrange



Leonhard Euler (1707-1783)



Joseph-Louis Lagrange (1736-1813)

- Born in Basel, Switzerland
- One of the most important mathematicians in history (Euler angle, Euler number, Euler formula,…)
- 13 children
- Published almost 900 papers and books, most of them in the last 20 years of his life
- Became totally blind in 1771

- Born in Turin, Italy
- One of 11 children, but only two of them lived to adulthood
- Autodidact
- Became professor in Turin at the age of 19
- Later he worked in Berlin and Paris
- Worked together with Euler on the calculus of variation

Gâteaux derivative

- A functional maps each element u of a vector space (e.g. a function) to a scalar $E(u)$.

- The Gâteaux derivative generalizes the directional derivative to infinite-dimensional vector spaces

$$\boxed{\left.\frac{\partial E(u)}{\partial u}\right|_h = \lim_{\epsilon \to 0} \frac{E(u+\epsilon h) - E(u)}{\epsilon} = \left.\frac{dE(u+\epsilon h)}{d\epsilon}\right|_{\epsilon=0}}$$

- This can be interpreted as the projection of the gradient to the direction h

$$\left.\frac{\partial E(u)}{\partial u}\right|_h = \left\langle \frac{dE(u)}{du}, h \right\rangle = \int \frac{dE(u)}{du}(\mathbf{x}) h(\mathbf{x}) d\mathbf{x}$$

- This gradient $\frac{dE(u)}{du}(\mathbf{x})$ is needed to minimize $E(u)$

Denoising example

- A continuous denoising energy is defined by the functional

$$E(u(\mathbf{x})) = \int_\Omega (u(\mathbf{x}) - I(\mathbf{x}))^2 + \alpha |\nabla u(\mathbf{x})|^2 \, d\mathbf{x}$$

- Rather than on a vector, the optimization problem is on a function

- Short writing: usually one does not explicitly indicate the dependency of the functions on $\mathbf{x}$

$$E(u) = \int_\Omega (u - I)^2 + \alpha |\nabla u|^2 \, d\mathbf{x}$$

- Necessary condition for a minimum:

$$\left.\frac{\partial E(u)}{\partial u}\right|_h = 0 \quad \forall h$$

Deriving the Euler-Lagrange equation

- Compute the Gâteaux derivative and simplify

$$
\begin{aligned}
\left.\frac{\partial E(u)}{\partial u}\right|_h &= \lim_{\epsilon\to 0}\frac{E(u+\epsilon h)-E(u)}{\epsilon} = \left.\frac{dE(u+\epsilon h)}{d\epsilon}\right|_{\epsilon=0} \\
&= \left.\frac{d}{d\epsilon}\int_\Omega (u+\epsilon h-I)^2+\alpha|\nabla(u+\epsilon h)|^2\,d\mathbf{x}\right|_{\epsilon=0} \\
&= \left.\frac{d}{d\epsilon}\int_\Omega (u+\epsilon h-I)^2+\alpha\left(\frac{d}{dx}(u+\epsilon h)\right)^2+\alpha\left(\frac{d}{dy}(u+\epsilon h)\right)^2\,d\mathbf{x}\right|_{\epsilon=0} \\
&= \left.\int_\Omega 2(u+\epsilon h-I)h+2\alpha\left(\frac{d}{dx}(u+\epsilon h)\frac{dh}{dx}+\frac{d}{dy}(u+\epsilon h)\frac{dh}{dy}\right)\,d\mathbf{x}\right|_{\epsilon=0} \\
&= \int_\Omega 2(u-I)h+2\alpha\left(u_x h_x+u_y h_y\right)\,d\mathbf{x}
\end{aligned}
$$

- Partial integration to turn h_x and h_y into h

$$
\begin{aligned}
\left.\frac{\partial E(u)}{\partial u}\right|_h &= \int_\Omega 2(u-I)h-2\alpha(u_{xx}h+u_{yy}h)\,d\mathbf{x}+(u_x h)_{\partial\Omega_x}+(u_y h)_{\partial\Omega_y} \\
&= \int_\Omega (2(u-I)-2\alpha(u_{xx}+u_{yy}))h\,d\mathbf{x}+((\mathbf{n}^\top\nabla u)h)_{\partial\Omega}
\end{aligned}
$$

where $\partial\Omega$ denotes the boundary and $\mathbf{n}$ the boundary normal.

Gradient and boundary conditions

In a minimum, the directional derivative must be 0 for all directions

$$\left.\frac{\partial E(u)}{\partial u}\right|_h = \int_\Omega (2(u-I) - 2\alpha(u_{xx}+u_{yy}))h\,d\mathbf{x} + ((\mathbf{n}^\top \nabla u)h)_{\partial\Omega} = 0$$

Yields two conditions:

1. The gradient must be zero → Euler-Lagrange equation:

$$\frac{dE(u)}{du} = 2(u-I) - 2\alpha(u_{xx}+u_{yy}) = 0$$

2. Boundary conditions:

$$(\mathbf{n}^\top \nabla u)_{\partial\Omega} = 0$$

These boundary conditions are called **natural boundary conditions** as they naturally emanate from the Gâteaux derivative. In the special case here, they coincide with **Neumann boundary conditions**.

Implementation

- Discretization of

$$\frac{dE(u)}{du} = (u - I) - \alpha(u_{xx} + u_{yy}) = 0$$

leads to a (in this case) linear system of equations

$$\frac{\partial E}{\partial u_{i,j}} \;=\; (u_{i,j} - I_{i,j}) - \alpha(u_{i-1,j} + u_{i,j-1} - 4u_{i,j} + u_{i+1,j} + u_{i,j+1}) = 0$$

- This is the same linear system one has to solve for minimizing the discrete energy from last class.

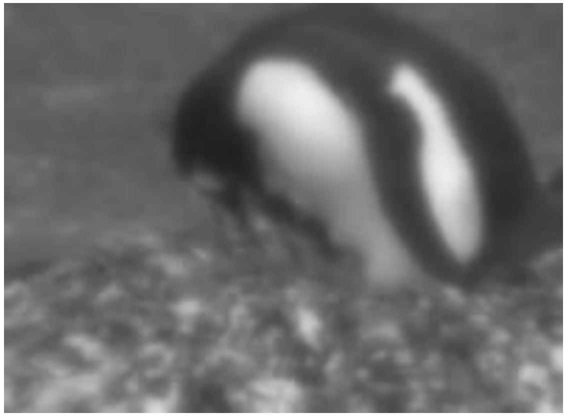
- Same linear solvers can be applied

- The two different discretization stages do not always lead to the same algorithm

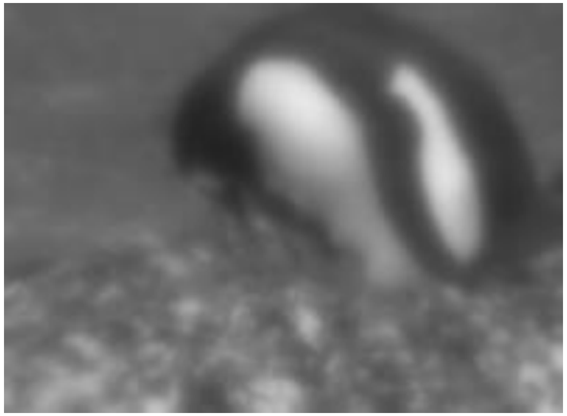
Smoothing results – do they remind you of some other method?

Input image

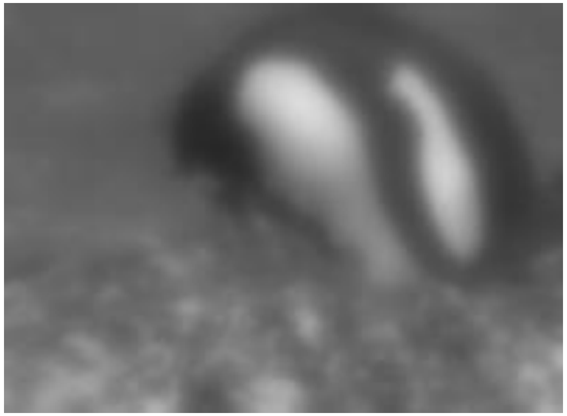
$\alpha = 1$



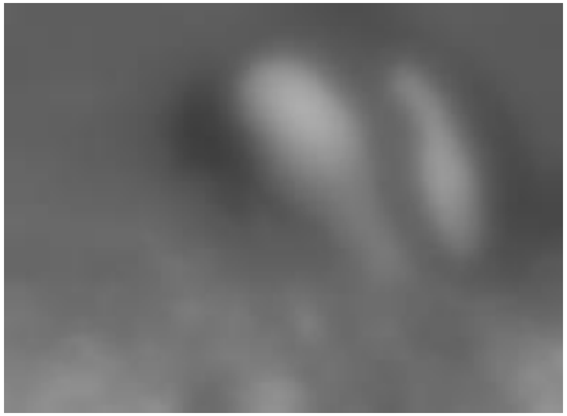
$\alpha = 5$



$\alpha = 10$

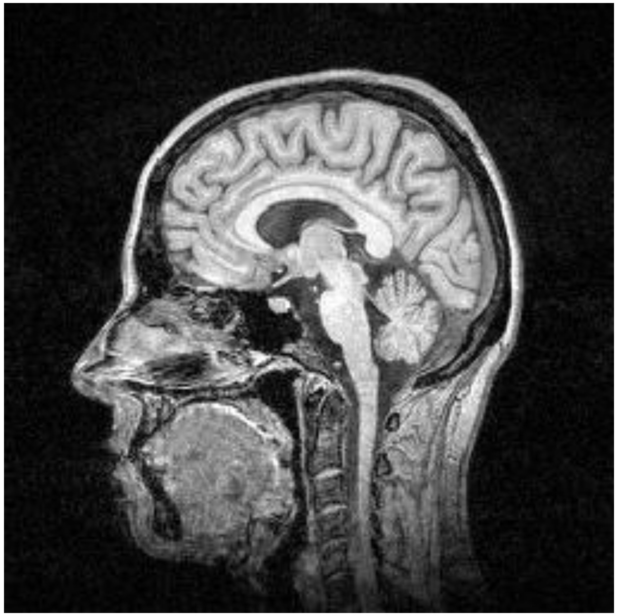


$\alpha = 20$

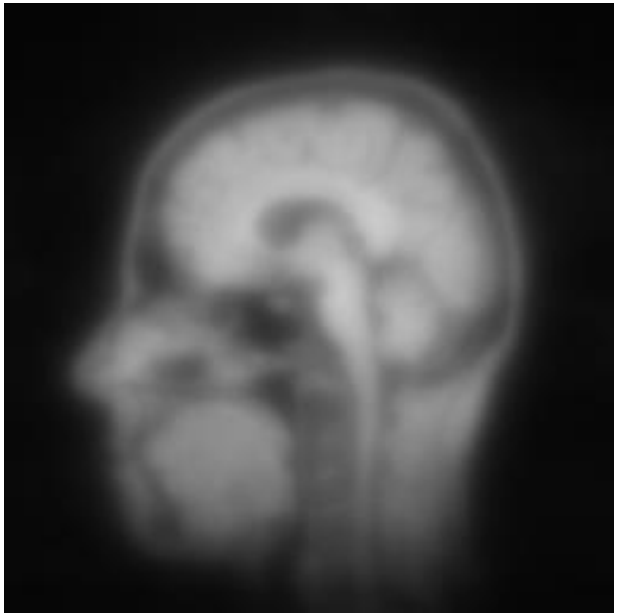


$\alpha = 100$

Tikhonov regularization vs. Gaussian smoothing

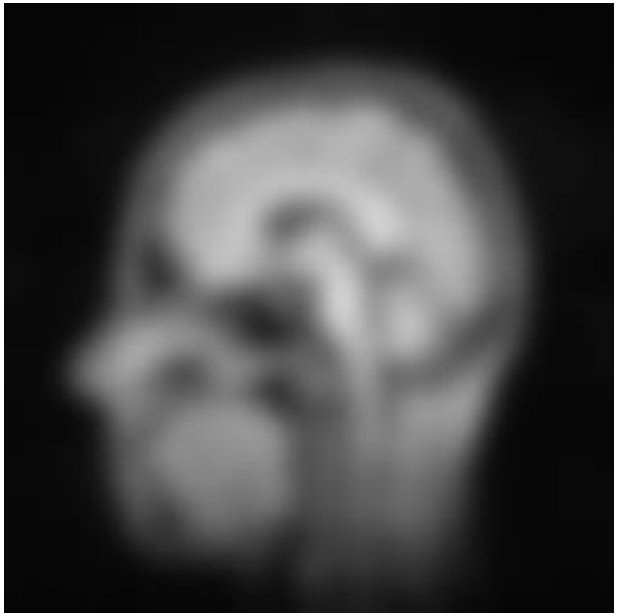


Input image



Tikhonov regularization

$\alpha = 20$



Gaussian smoothing

$\sigma^2 = 40$

Discontinuity preserving regularization

- A quadratic regularizer leads to blurred edges

- Why? Well, we said the result should be smooth everywhere, right?

- Not true for edges
  → we must allow for some exceptions (so-called **outliers**)

- This can be done by using non-quadratic regularizers

- Two ways to understand this (leading to the same solution)
  1. Nonlinear diffusion methods (Computer Vision course)
  2. Statistical interpretation of regularization

Statistical interpretation – the Bayesian approach

- Energy minimization can also emerge from a probabilistic approach taking into account likelihoods and prior probablities

- It builds upon the **Bayes formula:**

$$p(u|I) = \frac{p(I|u)p(u)}{p(I)}$$

a posteriori probability	evidence/likelihood
marginal probability	a priori probability

- Find solution u that maximizes the posterior probability

$p(u|I) \rightarrow$ max

→ maximum a-posteriori (MAP) approach

- Marginal does not depend on u → can be ignored in the maximization

$$p(u|I) \propto p(I|u)p(u)$$

How is this related to energy minimization?

- MAP estimation

$$p(u|I) \propto p(I|u)p(u) \quad \rightarrow \quad \text{max}$$

- Noise model for the data: i.i.d. Gaussian noise
(i.i.d. = independently and identically distributed)

$$p(I|u) \propto \prod_{\mathbf{x}\in\Omega} \exp\left(-(u(\mathbf{x}) - I(\mathbf{x}))^2\right)$$

- A-priori model: gradient is likely to be small (smoothness)
→ i.i.d. zero mean Gaussian noise with variance $1/2\alpha$ on the gradient

$$p(u) \propto \prod_{\mathbf{x}\in\Omega} \exp\left(-\frac{|\nabla u(\mathbf{x})|^2}{1/\alpha}\right)$$

- Turn maximization into minimization of the negative logarithm (which is monotonous) → energy minimization problem

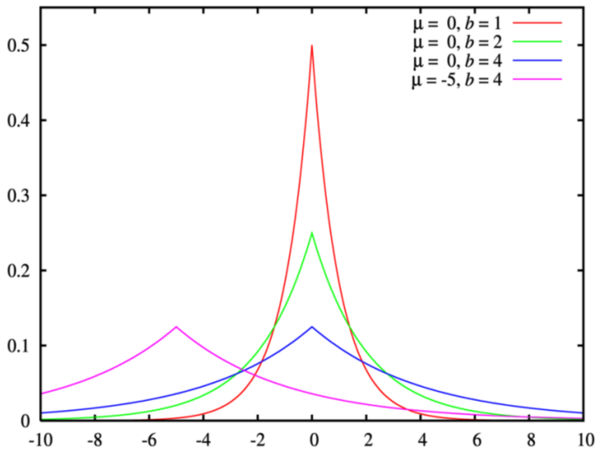
$$-\log p(I|u) - \log p(u) = \int_\Omega (u-I)^2 + \alpha|\nabla u|^2 d\mathbf{x} \quad \rightarrow \quad \text{min}$$

Statistical interpretation

- Other noise models lead to different penalizers in the energy functional

- Expecting outliers in the smoothness assumption, the Gaussian noise model should be replaced, e.g., by a Laplace distribution

$$p(u) \propto \prod_{\mathbf{x}\in\Omega} \exp\left(-\frac{|\nabla u(\mathbf{x})|}{1/\alpha}\right)$$



Laplace distribution

- Leads to an energy functional that preserves image edges

$$E(u) = \int_\Omega (u-I)^2 + \alpha|\nabla u| d\mathbf{x}$$

- More general:

$$E(u) = \int_\Omega (u-I)^2 + \alpha\Psi(|\nabla u|^2) d\mathbf{x}$$

here: $\Psi(s^2) = \sqrt{s^2}$

- (Quadratic) Tikhonov regularizer is another special case: $\Psi(s^2) = s^2$

How does this affect minimization?

- Corresponding Euler-Lagrange equation (verify at home):

$$\operatorname{div}\left(\Psi'\left(|\nabla u|^2\right)\nabla u\right) - \frac{u-I}{\alpha} = 0$$

where $\Psi'(s^2)$ is the derivative of $\Psi(s^2)$ with respect to its argument.

- So we get

$$\Psi(s^2) = \sqrt{s^2} \quad \Rightarrow \quad \Psi'(s^2) = \frac{1}{2\sqrt{s^2}}$$

$$\operatorname{div}\left(\frac{\nabla u}{2|\nabla u|}\right) - \frac{u-I}{\alpha} = 0$$

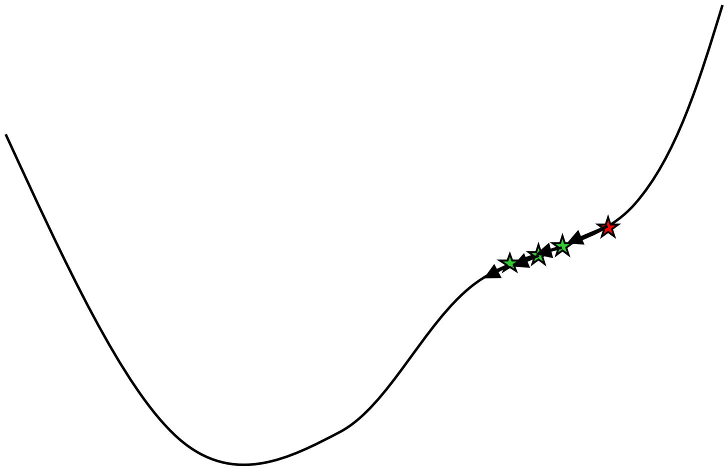
- Do you see the trouble?

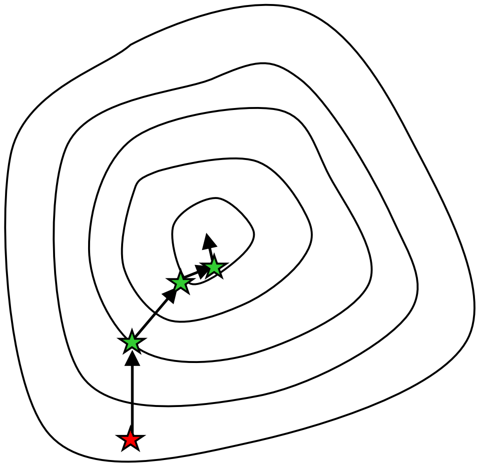
Gradient descent

- The Euler-Lagrange equation is nonlinear in the unknowns
  → discretization leads to a nonlinear system of equations

- A general way to minimize discrete or continuous energies (even if the gradient is nonlinear) is by **gradient descent**

- Iterative technique: start with an initial value, iteratively move in direction of largest decrease in energy (negative gradient direction)





- Converges to the next (with regard to the initialization) local minimum

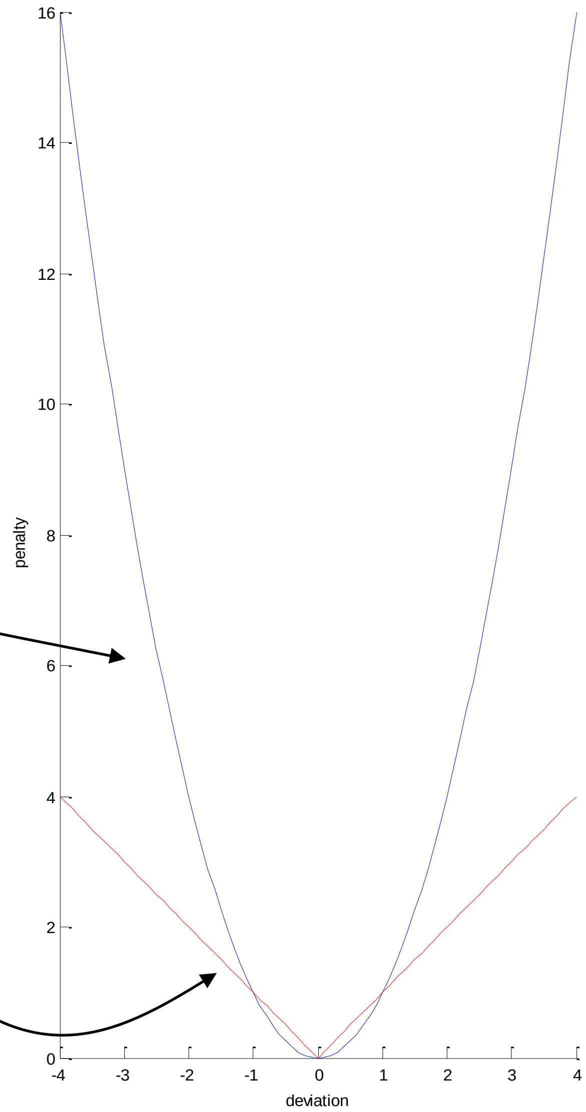
Is the problem convex or non-convex?

- Let’s see….

$$E(u) = \int_\Omega (u - I)^2 + \alpha|\nabla u| dx$$

clearly convex

This was close! But yes, convex



- Linear combination is convex → gradient descent will find a global optimum

Gradient descent

- Initialize u^0
- Introduce artificial time

$$\frac{\partial u}{\partial t} = -\frac{dE(u)}{du} = \mathrm{div}\left(\frac{\nabla u}{2|\nabla u|}\right) - \frac{u-I}{\alpha}$$

- Compute solution for $t \to \infty$
- Discrete steps of size τ forward in time

$$u^{k+1} = u^k + \tau\left(\mathrm{div}\left(\frac{\nabla u^k}{2|\nabla u^k|}\right) - \frac{u^k - I}{\alpha}\right)$$

- Will converge, if τ is "small enough" (more in course Computer Vision)
- Set $\Psi' = \frac{1}{\sqrt{|\nabla u|^2 + \epsilon^2}}$, $\epsilon = 0.01$, then $\tau \leq \frac{\epsilon}{4}$

→ slow convergence

A faster numerical scheme: lagged diffusivity

$$\text{div}\left(\Psi'\left(|\nabla u|^2\right)\nabla u\right) - \frac{u-I}{\alpha} = 0$$

- Keep the nonlinear prefactor fixed (now we have again a linear system) and compute updates in an iterative manner

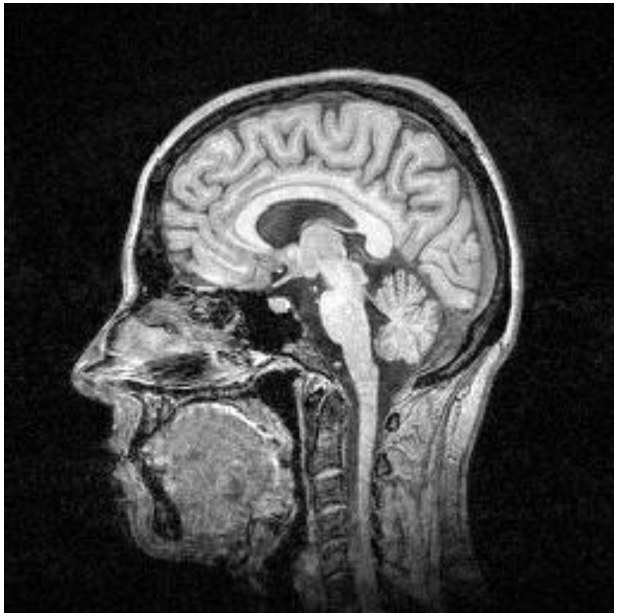
$$\text{div}\left(\Psi'^k \nabla u^{k+1}\right) - \frac{u^{k+1}-I}{\alpha} = 0$$

- This scheme is called **lagged diffusivity** and has been proven to converge if the linear system in each step is solved exactly

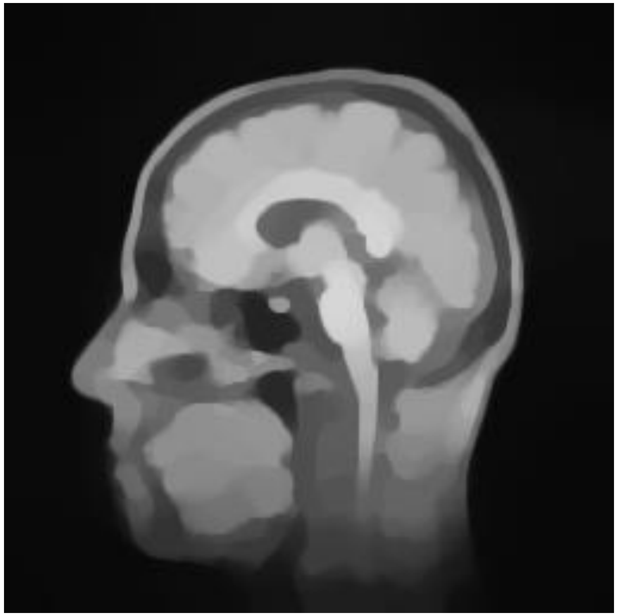
- In practice, only few iterations of the iterative linear solver are computed before Ψ' is updated → increased efficiency

- Much faster than gradient descent

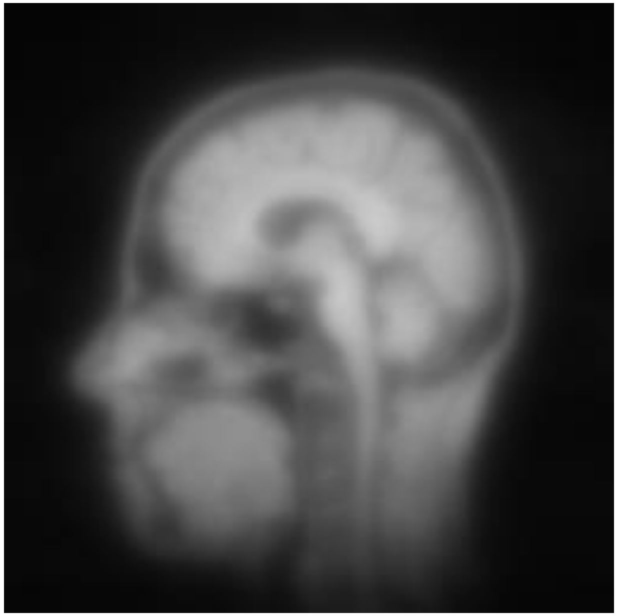
TV regularization vs. Tikhonov regularization



Input image



TV regularization
$\alpha = 20$



Tikhonov regularization
$\alpha = 20$

Summary

- The energy minimization framework also works with continuous models.
- Energy functions then turn into energy functionals where the unknowns are infinite-dimensional.
- Gradients of such functionals can be derived from the calculus of variation and the Gâteaux derivative. They lead to the Euler-Lagrange equation(s).
- The Bayes formula and the MAP approach lead to a statistical interpretation of many energy minimization problems
- We can design an edge-preserving smoothing method by choosing certain non-quadratic penalizers
- Optimization by gradient descent or the lagged diffusivity approach

Literature

- L. D. Elsgolc: Calculus of Variations, Pergamon Press, 1961. Reprint by Dover 2007.

- O. Scherzer, J. Weickert: Relations between regularization and diffusion filtering, Journal of Mathematical Imaging and Vision, 12:43-63, 2000.