Total Variation and the ROF image denoising problem

The continuous case and the derivation of the PDE associated to ROF problem

Andersen Ang

Department of Combinatorics and Optimization,
University of Waterloo, Waterloo, Canada

ms**x**ang@uwaterloo.ca, where $\mathbf{x} = \lfloor \pi \rfloor$
Homepage: angms.science

First draft: February 10, 2022 Last update: February 18, 2022

2D image in infinite dimensional space

Treat an observed image  as the discretization of a continuous two-variable function $u_0(x, y)$.

- $u_0(x, y) : \Omega \to \mathbb{R}$, x, y: coordinate of the image pixel.
- $\operatorname{dom} u_0 =: \Omega$ is a bounded closed subset of $\mathbb{R}^2$.
- Rudin-Osher-Fatemi[1] problem: we can recover a clean image u from u_0 by solving

$$\min_u \int_\Omega \|\nabla u\| dxdy \text{ st } \int_\Omega (u - u_0) dxdy = 0 \text{ and } \int_\Omega \frac{1}{2}\Big(u - u_0\Big)^2 dxdy = \sigma^2,$$

where $\|\cdot\|$ is the $\|\cdot\|_{L^2(\Omega)}$ norm.

[1] Rudin, Leonid I., Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms." Physica D: nonlinear phenomena 60, no. 1-4 (1992): 259-268.

On $\|\nabla u\|$

- Recall Multi-variable calculus 101: $u(x, y) : \Omega \to \mathbb{R}$ is a scalar field in $\mathbb{R}^2$.
- 2D gradient (vector) operator $\nabla = \dfrac{\partial}{\partial x}\mathbf{i} + \dfrac{\partial}{\partial y}\mathbf{j}$.
- Applying gradient operator on u gives a vector field

$$\nabla u(x, y) = \frac{\partial u(x, y)}{\partial x}\mathbf{i} + \frac{\partial u(x, y)}{\partial y}\mathbf{j}.$$

  i.e., every point $\nabla u(x, y)$ is a 2D vector.
- The norm of such vector $\|\nabla u\|$ tells the amount of variation of u at the point (x, y).
- By Pythagoras theorem, the norm is

$$\|\nabla u(x, y)\| = \sqrt{\Big(\frac{\partial u(x, y)}{\partial x}\Big)^2 + \Big(\frac{\partial u(x, y)}{\partial y}\Big)^2}.$$

Total variation

- We now know that $\|\nabla u(x,y)\|$ tells the amount of variation of the function u at the point (x,y) as

$$\|\nabla u(x,y)\| = \sqrt{\Big(\frac{\partial u(x,y)}{\partial x}\Big)^2 + \Big(\frac{\partial u(x,y)}{\partial y}\Big)^2}.$$

- Summing (integrate) $\|\nabla u(x,y)\|$ across all $(x,y) \in \Omega$ gives us the total variation within Ω

$$\|u\|_{\mathrm{BV}(\Omega)} := \int_\Omega \|\nabla u(x,y)\| \, dxdy = \int_\Omega \sqrt{\Big(\frac{\partial u(x,y)}{\partial x}\Big)^2 + \Big(\frac{\partial u(x,y)}{\partial y}\Big)^2} \, dxdy,$$

where BV(Ω) denotes bounded variation within the domain Ω.

- Shorthand notation: u denotes $u(x,y)$ and u_x denotes $\dfrac{\partial u(x,y)}{\partial x}$.

Simple examples to understand TV

- Single-variable case $\|u\|_{\mathrm{BV}(\Omega)} = \int_\Omega \sqrt{\left(\frac{du(x)}{dx}\right)^2}\, dx$ at the interval $\Omega = [0, 1]$.

- $\sqrt{\left(\frac{du(x)}{dx}\right)^2} = \left|\frac{du(x)}{dx}\right|$ because $\sqrt{x^2} = |x|$.

- Constant function $u(x) = c$: $\quad \frac{du}{dx} = 0 \implies \int_0^1 0\, dx = 0.$

  Comment: $u(x)$ is constant so no variation, thus $\|u\|_{\mathrm{BV}(\Omega)} = 0$.

- For linear function $u(x) = mx + c$, $\quad \frac{du}{dx} = m \implies \int_0^1 m\, dx = m.$

  Comment: this is exactly the amount of change for u in $\Omega = [0, 1]$.

ROF Image denoising problem

- Additive white noise model: $u_0(x, y) = u(x, y) + \eta(x, y).$ (1)
  - $u_0(x, y) : \Omega \to \mathbb{R}$ observed (noisy) image
  - $u(x, y) : \Omega \to \mathbb{R}$ desired (clean) image
  - $\eta(x, y) : \Omega \to \mathbb{R}$ additive white noise
    - Assumed to be zero mean.
    - Assumed to have standard deviation σ / variance σ^2.
- Obtain $u(x, y)$ by solving the ROF[2] minimization problem

$$\min_u \|u\|_{\mathrm{BV}(\Omega)} \text{ st } \int_\Omega (u - u_0)\, dxdy = 0 \text{ and } \int_\Omega \frac{1}{2} (u - u_0)^2\, dxdy = \sigma^2.$$

  The ROF paper states
  - the 1st constraint means η has zero mean
  - the 2nd constraint means η has variance equal to σ^2.

---

[2] Rudin, Leonid I., Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms." Physica D: nonlinear phenomena 60, no. 1-4 (1992): 259-268.

Regularization form of ROF problem

- Original ROF problem

$$\min_u \|u\|_{\mathrm{BV}(\Omega)} \;\text{ s.t. } \int_\Omega (u-u_0)dxdy = 0 \;\text{ and }\; \int_\Omega \frac{1}{2}(u-u_0)^2 dxdy = \sigma^2.$$

- Consider a regularization expression (penalty form)

$$\min_u \|u\|_{\mathrm{BV}(\Omega)} + \lambda\left(\int_\Omega \frac{1}{2}(u-u_0)^2 dxdy - \sigma^2\right) \quad \text{s.t. } \int_\Omega (u-u_0)dxdy = 0$$

where $\lambda \geq 0$ is a regularization parameter.

- Later we will see that constraint $\int_\Omega (u-u_0)dxdy = 0$ can be enforced algorithmically by boundary condition.

Image denoising is a (strongly) convex optimization problem

- Ignoring constant term, the regularized ROF problem becomes

$$\min_u \|u\|_{\mathrm{BV}(\Omega)} + \underbrace{\frac{\lambda}{2}\int_\Omega (u-u_0)^2 dxdy}_{g(u):=\|u-u_0\|^2_{L^2(\Omega)}} \quad \text{s.t.} \quad \int_\Omega (u-u_0)dxdy = 0.$$

- Let $f(u) = \|u\|_{\mathrm{BV}(\Omega)}$ and $g(u) = \frac{\lambda}{2}\|u-u_0\|^2_{L^2(\Omega)}$, the problem becomes

$$(\mathcal{P}) \; : \; \min_u f(u) + g(u) \quad \text{s.t.} \quad \int_\Omega (u-u_0)dxdy = 0.$$

- f and g are convex and $f+g$ is strongly convex.
  The domain of the optimization variable is a closed convex set.
  Hence $\mathcal{P}$ is a convex optimization problem and it has a global unique solution u^*_λ.

1st-order optimality condition

$$\min_u \underbrace{\|u\|_{\mathrm{BV}(\Omega)}}_{f(u)} + \underbrace{\frac{\lambda}{2}\|u-u_0\|^2_{L^2(\Omega)}}_{g(u)} \quad \text{s.t.} \quad \int_\Omega (u-u_0)dxdy = 0.$$

- By 1st-order optimality condition, u^*_λ is a minimizer of the problem iff

$$0 \in \partial_u f(u^*_\lambda) + \partial_u g(u^*_\lambda),$$

where the set of subdifferential $\partial_u f(u^*_\lambda) := \{q \mid f(u) \geq f(u^*_\lambda) + \langle q, u - u^*_\lambda\rangle\}$.

- Note that we are now in infinite dimensional space, so the inner product is refer to the inner product in infinite dimensional space: $\langle f(x,y), g(x,y)\rangle = \int_\Omega f(x,y)g(x,y)dxdy$.

Subgradient of $g(u) = \frac{\lambda}{2}\|u - u_0\|^2_{L^2(\Omega)}$

- For $g(u) = \frac{\lambda}{2}\|u - u_0\|^2_{L^2(\Omega)}$,
$$\partial_u g(u) = \lambda(u - u_0).$$

- (Generalization) For $g(u) = \frac{\lambda}{2}\|Ku - u_0\|^2_{L^2(\Omega)}$, where $K : L^2(\Omega) \to \mathcal{H}$ with $\mathcal{H}$ being a Hilbert space, i.e., Ku denotes $K(u)$ and
$$g(u) = \int_\Omega K\big(u(x, y)\big) - u_0(x, y)dxdy,$$
we have
$$\partial_u g(u) = \lambda K^*(Ku - u_0),$$
where K^* is the Hermitian transpose of K.

Subgradient of $f(u) = \|u\|_{\mathrm{BV}(\Omega)}$

- For $f(u) = \|u\|_{\mathrm{BV}(\Omega)} = \int_\Omega \|\nabla u\| \, dxdy$, we have $\partial_u f(u) = -\mathrm{div}\left(\frac{\nabla u}{\|\nabla u\|}\right) = -\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|}\right), \quad \|\nabla u\| \neq 0.$.
  Note that it is undefined at (x, y) where $\|\nabla u(x, y)\| = 0$.
- Recall vector field $\nabla u = \frac{\partial u}{\partial x}\mathbf{i} + \frac{\partial u}{\partial y}\mathbf{j}$ and $\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j}$, thus

$$
\begin{aligned}
\partial_u f(u) &= -\Big(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial x}\mathbf{j}\Big) \cdot \left(\frac{\frac{\partial u}{\partial x}\mathbf{i} + \frac{\partial u}{\partial y}\mathbf{j}}{\sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2}} \right), \quad \|\nabla u\| \neq 0. \\
&= -\frac{\partial}{\partial x}\left(\frac{\frac{\partial u}{\partial x}}{\sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2}} \right) - \frac{\partial}{\partial x}\left(\frac{\frac{\partial u}{\partial y}}{\sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2}} \right), \quad \|\nabla u\| \neq 0.
\end{aligned}
$$

- This is wrong:

$$
-\Big(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial x}\mathbf{j}\Big) \cdot \left(\frac{\frac{\partial u}{\partial x}\mathbf{i} + \frac{\partial u}{\partial y}\mathbf{j}}{\sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2}} \right) = -\frac{\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}}{\sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2}}. \qquad \text{(Wrong)}
$$

$\partial_u f(u)$ is a nonlinear Laplacian

- Vector calculus fact: divergence of gradient is Laplacian

$$\Delta u = \nabla^2 u = \nabla \cdot (\nabla u)$$

- Therefore, if $\|\nabla u\| = \sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2} = 1$, then we have Laplacian

$$\partial_u f(u) = -\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|} \right) = -\nabla \cdot \nabla u = -\nabla^2 u.$$

- In general, $\|\nabla u\| \neq 1$ thus $\partial_u f(u)$ is a nonlinear Laplacian

$$\begin{aligned} \partial_u f(u) &= -\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|} \right) \\ &= -\frac{\|\nabla u\| \nabla \cdot \nabla u - \nabla \|\nabla u\| \cdot \nabla u}{\|\nabla u\|^2} \\ &= -\frac{\|\nabla u\| \nabla^2 u - (\nabla \|\nabla u\|) \cdot \nabla u}{\|\nabla u\|^2}. \end{aligned}$$

A fast review of Calculus of variation

- **Definition** A **functional** is a mapping $F : X \to \mathbb{R}$ where the set $X = \mathrm{dom}\, f$ is a infinite dimensional function space.
  i.e., a functional = a function of function.

- We focus on X being a Hilbert space[3].

- **Definition** A **linear** functional on Hilbert space X is a mapping $\ell : X \to \mathbb{R}$ such that $(\forall a, b \in \mathbb{R})(\forall \mathbf{x} \in X)(\forall \mathbf{y} \in X)$ we have $\ell(a\mathbf{x} + b\mathbf{y}) = a\ell(\mathbf{x}) + b\ell(\mathbf{y})$.

- **Definition** A functional $\ell(\mathbf{x})$ is **bounded** if $(\forall \mathbf{x} \in X)\ \exists c \text{ s.t. } |\ell(\mathbf{x})| \leq c\|\mathbf{x}\|$.

- **Theorem**[4] Let X be a Hilbert space. For all bounded linear functional $\ell : X \to \mathbb{R}$, there exists a (unique) vector $\mathbf{z} \in X$ such that $(\forall \mathbf{x} \in X)\ \ \ell(\mathbf{x}) = \langle \mathbf{z}, \mathbf{x} \rangle$.

---

[3] = a complete inner product space.

[4] Riesz representation theorem.

- **Definition** Let $F : X \to \mathbb{R}$ be a functional on a Hilbert space X. If the limit

$$F'(\mathbf{x};\mathbf{v}) = \lim_{h\to 0^+} \frac{F(\mathbf{x}+h\mathbf{v}) - F(\mathbf{x})}{h} = \frac{d}{dh}F(\mathbf{x}+h\mathbf{v})\Big|_{h=0}$$

exists, it is called the **directional derivative** of F at $\mathbf{x}$ in the direction $\mathbf{v}$.
  - **Definition** Furthermore, we call F is **Gâteaux differentiable** if $F'(\mathbf{x};\mathbf{v})$ is a bounded linear functional of $\mathbf{v}$.
  - By theorem, $\exists \mathbf{u}_\mathbf{x} \in X$ s.t. $(\forall \mathbf{x} \in X)\ \ F'(\mathbf{x};\mathbf{v}) = \langle \mathbf{u}_\mathbf{x}, \mathbf{v}\rangle$. The function (not vector!) $\mathbf{u}_\mathbf{x}$ is called the **Gateaux derivative** of F at the function (not vector!) $\mathbf{x}$.
- Now recall $\|u\|_{\mathrm{BV}}(\Omega) = \displaystyle\int_\Omega \|\nabla u(x,y)\| dxdy$ and we want to find $\partial\|u\|_{\mathrm{BV}}(\Omega)$. We can compute it by Gateaux derivative, but it will be tedious. A short cut is to use the Euler-Lagrangian equation.
- **Lemma** For $F(u) = \displaystyle\int_\Omega L(\mathbf{x}, u(\mathbf{x}), \nabla u(\mathbf{x}))d\mathbf{x}$ with a a Lagrangian function L that $\mathbf{x} = [x_1, \dots, x_n]$ and $u : X \to \mathbb{R}$. Let $\mathbf{w} = [w_1, \dots, w_n] \in \mathbb{R}^n$, $w_i = [\nabla u(\mathbf{x})]_i$. Based on the same technique that derives the Euler-Lagrangian equation, the Gateaux derivative $F'(u)$ can be written as

$$F'(u) = \frac{\partial L}{\partial u}(\mathbf{x}, u, \nabla u) - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\Big(\frac{\partial L}{\partial w_i}(\mathbf{x}, u, \nabla u)\Big).$$

Prove $\partial \|u\|_{\mathrm{BV}(\Omega)} = -\mathrm{div}\left(\frac{\nabla u}{\|\nabla u\|}\right)$

- $F(u) = \int_\Omega \|\nabla u(x,y)\| dxdy$
- Let $L(\mathbf{x}, u, \mathbf{w}) = \sqrt{w_1^2 + w_2^2}$, where $w_i = [\nabla u]_i$. I.e., $w_1 = u_x$ and $w_2 = u_y$.
- Partial derivatives of $L(\mathbf{x}, u, \mathbf{w})$:

$$\frac{\partial L}{\partial u}(\mathbf{x}, u, \mathbf{w}) = 0, \quad \frac{\partial L}{\partial w_1}(\mathbf{x}, u, \mathbf{w}) = \frac{w_1}{\sqrt{w_1^2 + w_2^2}}, \quad \frac{\partial L}{\partial w_2}(\mathbf{x}, u, \mathbf{w}) = \frac{w_2}{\sqrt{w_1^2 + w_2^2}}.$$

- Gateaux derivative

$$\begin{aligned} F'(u) &= \underbrace{\frac{\partial L}{\partial u}(\mathbf{x}, u, \mathbf{w})}_{=0} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\frac{\partial L}{\partial w_i}(\mathbf{x}, u, \mathbf{w})\right) \\ &= -\frac{\partial}{\partial x}\left(\frac{w_1}{\sqrt{w_1^2 + w_2^2}}\right) - \frac{\partial}{\partial y}\left(\frac{w_1}{\sqrt{w_1^2 + w_2^2}}\right) \\ &= -\left[\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j}\right] \cdot \left(\frac{u_x\mathbf{i} + u_y\mathbf{j}}{\sqrt{u_x^2 + u_y^2}}\right) \qquad = -\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|}\right) = -\mathrm{div}\left(\frac{\nabla u}{\|\nabla u\|}\right). \end{aligned}$$

1st-order optimality condition and Euler-Lagrange equation

- ROF image denoising minimization problem

$$\min_u \|u\|_{\mathrm{BV}(\Omega)} + \frac{\lambda}{2}\|u-u_0\|^2_{L^2(\Omega)} \quad \text{s.t.} \quad \int_\Omega (u-u_0)dxdy = 0.$$

- Subgradient 1st-order optimality condition of the cost function

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

- We get a similar expression of optimality based on Euler-Lagrange equation

$$0 = -\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|}\right) + \lambda(u-u_0), \ \|\nabla u\| \neq 0.$$

- Here actually we can see that subgradient optimality condition is slightly more general than Euler-Lagrange equation.

The Euler-Lagrange equation viewpoint of 1st-order optimality condition

- The ROF problem in integral form

$$\min_u \int_\Omega \underbrace{\|\nabla u\| + \frac{\lambda}{2}(u-u_0)^2}_{L} dxdy \quad \text{s.t.} \quad \int_\Omega (u-u_0)dxdy = 0.$$

- By the Fundamental lemma of Calculus of variation, the Euler-Lagrange equation of L has to be zero (after applying the gradient operator ∇), hence

$$0 \quad = \quad -\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|} \right) + \lambda(u-u_0), \;\; \|\nabla u\| \neq 0.$$

  where the negative sign comes from the second term in the Euler-Lagrange equation.

- Recall Euler-Lagrange equation: for $S(\mathbf{q}) = \displaystyle\int_\Omega L\big(t, \mathbf{q}(t), \dot{\mathbf{q}}(t)\big)dt$, function $\mathbf{q}$ is a stationary point of S if and only if

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} = 0.$$

ROF problem is equivalent to a nonlinear elliptic PDE

- We now see that, solving the ROF image denoising minimization problem

$$\min_u \|u\|_{\mathrm{BV}(\Omega)} + \frac{\lambda}{2}\|u - u_0\|^2_{L^2(\Omega)} \quad \text{s.t.} \quad \int_\Omega (u - u_0)dxdy = 0.$$

  is equivalent to solving the nonlinear elliptic partial differential equation

$$\begin{cases} \underbrace{-\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|}\right)}_{\text{nonlinear}} + \lambda(u - u_0) = 0 & u \in \Omega \\ \frac{\partial u}{\partial \mathbf{n}} = 0 & u \in \partial\Omega \end{cases}$$

- The constraint $\frac{\partial u}{\partial \mathbf{n}} = 0$ constraint is the Neumann boundary condition, it guarantees $\int_\Omega (u - u_0)dxdy = 0$ holds in the ROF problem.
- $\mathbf{n}$ is the normal vector to the boundary $\partial\Omega$.

Gradient descent

$$\underbrace{\min_{u} \|u\|_{\mathrm{BV}(\Omega)} + \frac{\lambda}{2}\|u - u_0\|^2_{L^2(\Omega)}}_{F} \text{ s.t. } \int_{\Omega} (u - u_0) dx dy = 0. \tag{$\dagger$}$$

- Recall the Gateaux derivative of the cost function in (†) is

$$-\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|} \right) + \lambda (u - u_0).$$

- So naturally have a gradient descent algorithm to solve (†): just evolve the gradient flow

$$\frac{\partial u}{\partial t} = -F'(u),$$

which is exactly the algorithm proposed in the ROF paper.

Last page - summary

- Total variation

$$\|u\|_{\mathrm{BV}(\Omega)} := \int_\Omega \|\nabla u(x,y)\| \, dxdy = \int_\Omega \sqrt{\left(\frac{\partial u(x,y)}{\partial x}\right)^2 + \left(\frac{\partial u(x,y)}{\partial y}\right)^2} \, dxdy,$$

- ROF image denoising problem (in regularization form)

$$\min_u \|u\|_{\mathrm{BV}(\Omega)} + \frac{\lambda}{2}\|u - u_0\|^2_{L^2(\Omega)} \quad \text{s.t.} \quad \int_\Omega (u - u_0) dxdy = 0.$$

- Equivalent to a nonlinear elliptic PDE

$$\begin{cases} \nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|}\right) + \lambda(u - u_0) = 0 & u \in \Omega \\ \dfrac{\partial u}{\partial \mathbf{n}} = 0 & u \in \partial\Omega \end{cases}$$

Next

- Finite approximation of continuous TV: discrete TV seminorm
- Application of discrete TV seminorm in image denoising
- Algorithms for solving discrete TV seminorm

End of document