MA3111: Mathematical Image Processing
Variational Image Denoising

Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 320317, Taiwan

First version: May 29, 2021/Last updated: June 25, 2025

Outline of “variational image denoising”

In this lecture, we will give a brief introduction to the topics:

- *The Rudin-Osher-Fatemi total variation model for image denoising.*
- *Calculus of variations: the Euler-Lagrange equation.*

The material of this lecture is based on

- P. Getreuer, Rudin-Osher-Fatemi total variation denoising using split Bregman, *Image Processing On Line,* 2 (2012), pp. 74-95.
- T. Goldstein and S. Osher, The split Bregman method for L^1 regularized problems, *SIAM Journal on Imaging Sciences,* 2 (2009), pp. 323-343.
- L. I. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D,* 60 (1992), pp. 259-268.

Total variation (總變差)

Let $\Omega := (a, b) \subset \mathbb{R}$ be an open bounded interval. Let $\mathcal{P}_n = \{x_0, x_1, \cdots, x_n\}$, with $x_0 = a$ and $x_n = b$, be an arbitrary partition of $\overline{\Omega} = [a, b]$ and $\Delta x_i = x_i - x_{i-1}$, for $i = 1, 2, \cdots, n$. The total variation of a real-valued function $u : \overline{\Omega} \to \mathbb{R}$ is defined as

$$\|u\|_{TV(\Omega)} := \sup_{\mathcal{P}_n} \sum_{i=1}^{n} \left| u(x_i) - u(x_{i-1}) \right|.$$

If $\|u\|_{TV(\Omega)} < \infty$, then we say that u is a function of bounded variation.

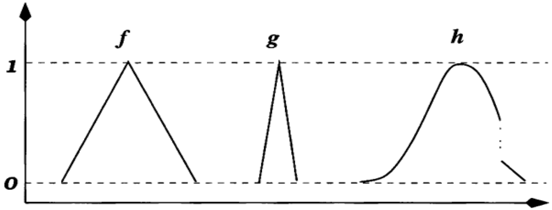
Remarks:

- *If u is a smooth function, then we have*

$$\|u\|_{TV(\Omega)} = \sup_{\mathcal{P}_n} \sum_{i=1}^{n} \left| \frac{u(x_i) - u(x_{i-1})}{\Delta x_i} \right| \Delta x_i = \int_{\Omega} |u'(x)| \, dx.$$

- $\|u\|_{TV(\Omega)} = 0$ *does not imply* $u \equiv 0$; *any constant function* u *has* $\|u\|_{TV(\Omega)} = 0 \implies \|u\|_{TV(\Omega)}$ *is not a norm on any vector space.*

Examples of bounded variation functions



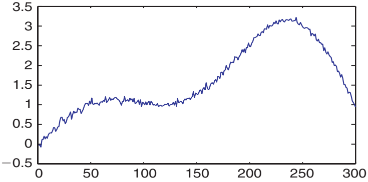
All these three functions f, g and h have total variation 2

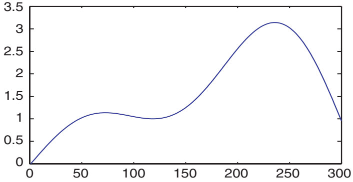
Remarks:

- *A function of bounded variation is not necessarily differentiable.*
- *Since we mainly work on digital images in discrete domains, we can tacitly assume the differentiability of u without loss of generality.*

Image denoising

- The total variation of $u = \|u\|_{TV(\Omega)} = \int_\Omega |u'(x)|\, dx$ if u is a smooth function.
- Image denoising is the problem of removing noise from a noisy image.
- *minimizing* $\left(\int_\Omega |u'(x)|\, dx + \underbrace{\textit{some data fidelity term}}_{e.g.,\ \int_\Omega (u(x)-f(x))^2 dx}\right) \Rightarrow$ *denoising!*





A noisy 1-D signal and its denoising version

The bounded variation space $BV(\Omega)$

Let Ω be an open subset of $\mathbb{R}^2$. The space of functions of bounded variation $BV(\Omega)$ is defined as the space of real-valued function $u \in L^1(\Omega)$ such that the total variation is finite, i.e.,

$$BV(\Omega) = \{u \in L^1(\Omega) : \|u\|_{TV(\Omega)} < \infty\},$$

where

- $\|u\|_{TV(\Omega)} = \sup \left\{ \int_\Omega u \nabla \cdot \varphi d\boldsymbol{x} : \varphi \in C_c^1(\Omega, \mathbb{R}^2), \|\varphi\|_{(L^\infty(\Omega))^2} \leq 1 \right\}$
- $C_c^1(\Omega, \mathbb{R}^2)$ is the space of continuously differentiable vector functions with compact support in Ω.
- $L^1(\Omega)$ and $L^\infty(\Omega)$ are the usual $L^p(\Omega)$ space for $p = 1$ and $p = \infty$, respectively, equipped with the $\|\cdot\|_{L^p(\Omega)}$ norm.
- *$BV(\Omega)$ is a Banach space with the norm,*

$$\|u\|_{BV(\Omega)} := \|u\|_{L^1(\Omega)} + \|u\|_{TV(\Omega)}.$$

The ROF total variation regularization model

Let $f : \overline{\Omega} \subset \mathbb{R}^2 \to \mathbb{R}$ be a given noisy image. Rudin, Osher, and Fatemi (*Physica D, 1992*) proposed the following TV/L2 model for image denoising:

$$\min_{u \in BV(\Omega) \cap L^2(\Omega)} \Bigg(\underbrace{\|u\|_{TV(\Omega)}}_{regularizer} + \frac{\lambda}{2} \underbrace{\int_\Omega \big(u(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2 \, d\boldsymbol{x}}_{data\ fidelity} \Bigg),$$

where $\lambda > 0$ is a tuning parameter which controls the regularization strength. Notice that

- A smaller value of λ will lead to a more regular solution.
- The space of functions with bounded variation help remove noise and preserve edges in the image.
- The *TV* term allows the solution to have discontinuities.

The existence, uniqueness and stability of solution

Theorem: *Consider the ROF total variation model. Then we have*

(1) *If u is smooth, then* $\|u\|_{TV(\Omega)} = \int_\Omega |\nabla u|\, d\boldsymbol{x} := \int_\Omega \sqrt{u_x^2 + u_y^2}\, d\boldsymbol{x}$.

(2) *If $f \in L^2(\Omega)$, then the minimizer exists and is unique and is stable in L^2 with respect to perturbations in f.*

ROF model for image denoising: Below we assume that *u is smooth,* and we denote the function vector space $BV(\Omega) \cap L^2(\Omega)$ as $\mathcal{V}$:

$$\min_{u \in \mathcal{V}} \left(\int_\Omega |\nabla u|\, d\boldsymbol{x} + \frac{\lambda}{2} \int_\Omega \big(u(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2\, d\boldsymbol{x} \right).$$

Let $E[\cdot]$ be *the energy functional* over the function vector space $\mathcal{V}$,

$$E[u] := \int_\Omega |\nabla u|\, d\boldsymbol{x} + \frac{\lambda}{2} \int_\Omega \big(u(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2\, d\boldsymbol{x}.$$

Calculus of variations (變分法)

- Calculus of variations is a branch of mathematical analysis that deals with *maximizing or minimizing functionals.* A real-valued *functional* is a mapping from a subset of function vector space $\mathcal{V}$ to the real numbers.

  (1) A real-valued function, e.g., $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$.

  (2) A real-valued functional, e.g., $E : S \subseteq \mathcal{V} \to \mathbb{R}$.

- Functionals are often expressed as definite integrals involving functions and their derivatives, e.g.,

$$E[v] := \int_\Omega L\big(x, y, v(x,y), v_x(x,y), v_y(x,y)\big)\, d\mathbf{x},$$

  for a given smooth function L.

  The interest is in extremal functions that make the functional attains a maximum or minimum value.

- *The extrema of functionals may be obtained by finding functions where the "functional derivative" is equal to zero. This leads to solving the associated Euler-Lagrange equation.*

Calculus of variations: a necessary condition

Let $\Omega \subset \mathbb{R}^2$ be an open bounded domain. We consider the following real-valued functional,

$$E[v] := \int_\Omega L(x, y, v(x,y), v_x(x,y), v_y(x,y))\, d\boldsymbol{x},$$

where we assume that $v \in C^2(\overline{\Omega})$ and $L \in C^2$ with respect to its arguments $\boldsymbol{x} = (x,y)$, v, v_x and v_y.

- If $E[v]$ attains a local minimum or maximum at u and $\eta(x,y)$ is a smooth function on $\overline{\Omega}$, then for ε close to 0, we have

  $$E[u] \le E[u + \varepsilon\eta], \quad (\text{or } E[u] \ge E[u + \varepsilon\eta])$$

  where $\delta u := \varepsilon\eta$ is called the variation of u.

- Define $\Phi(\varepsilon) := E[u + \varepsilon\eta]$ in the variable ε. Then we have

  $$0 = \Phi'(0) = \lim_{\varepsilon \to 0} \frac{E[u+\varepsilon\eta] - E[u]}{\varepsilon - 0} = \int_\Omega \frac{dL}{d\varepsilon}\bigg|_{\varepsilon=0} d\boldsymbol{x} = 0,$$

  which is just a necessary condition.

The total derivative of L

Taking the total derivative of $L(x, y, v, v_x, v_y)$, where $v = u + \varepsilon\eta$ $v_x = u_x + \varepsilon\eta_x$ and $v_y = u_y + \varepsilon\eta_y$, we have

$$
\begin{aligned}
\frac{dL}{d\varepsilon} &= \frac{\partial L}{\partial x}\frac{dx}{d\varepsilon} + \frac{\partial L}{\partial y}\frac{dy}{d\varepsilon} + \frac{\partial L}{\partial v}\frac{dv}{d\varepsilon} + \frac{\partial L}{\partial v_x}\frac{dv_x}{d\varepsilon} + \frac{\partial L}{\partial v_y}\frac{dv_y}{d\varepsilon} \\
&= \frac{\partial L}{\partial x}0 + \frac{\partial L}{\partial y}0 + \frac{\partial L}{\partial v}\eta + \frac{\partial L}{\partial v_x}\eta_x + \frac{\partial L}{\partial v_y}\eta_y \\
&= \frac{\partial L}{\partial v}\eta + \frac{\partial L}{\partial v_x}\eta_x + \frac{\partial L}{\partial v_y}\eta_y \\
&= \frac{\partial L}{\partial v}\eta + \left(\frac{\partial L}{\partial v_x}, \frac{\partial L}{\partial v_y}\right) \cdot \nabla\eta.
\end{aligned}
$$

The divergence theorem

Let $\Omega \subset \mathbb{R}^n$ be an open bounded domain with smooth boundary $\partial\Omega$. Let $A : \overline{\Omega} \subset \mathbb{R}^n \to \mathbb{R}^n$ be a smooth vector-valued function with $A(\boldsymbol{x}) = (A_1(\boldsymbol{x}), A_2(\boldsymbol{x}), \cdots, A_n(\boldsymbol{x}))^\top$. Then we have

$$\boxed{\int_\Omega \nabla \cdot A \, d\boldsymbol{x} = \int_{\partial\Omega} A \cdot \boldsymbol{n} \, d\sigma}$$

where

$$\begin{aligned} \nabla \cdot A(\boldsymbol{x}) &:= \frac{\partial A_1(\boldsymbol{x})}{\partial x_1} + \frac{\partial A_2(\boldsymbol{x})}{\partial x_2} + \cdots + \frac{\partial A_n(\boldsymbol{x})}{\partial x_n}, \\ \boldsymbol{n}(\boldsymbol{x}) &:= (n_1(\boldsymbol{x}), n_2(\boldsymbol{x}), \cdots, n_n(\boldsymbol{x}))^\top \\ &\quad \text{is the outward unit normal to } \partial\Omega. \end{aligned}$$

Application to two-dimensional domains ($n = 2$)

Assume $v : \overline{\Omega} \subset \mathbb{R}^2 \to \mathbb{R}$ and $w : \overline{\Omega} \subset \mathbb{R}^2 \to \mathbb{R}$ are smooth functions.

- If $A = (vw, 0)^\top$ then $\int_\Omega \frac{\partial(vw)}{\partial x}\, d\boldsymbol{x} = \int_{\partial\Omega} vwn_1\, d\boldsymbol{\sigma}$, which implies

$$\int_\Omega v \frac{\partial w}{\partial x}\, d\boldsymbol{x} + \int_\Omega \frac{\partial v}{\partial x} w\, d\boldsymbol{x} = \int_{\partial\Omega} vwn_1\, d\boldsymbol{\sigma}. \qquad (\star_1)$$

- If $A = (0, vw)^\top$ then $\int_\Omega \frac{\partial(vw)}{\partial y}\, d\boldsymbol{x} = \int_{\partial\Omega} vwn_2\, d\boldsymbol{\sigma}$, which implies

$$\int_\Omega v \frac{\partial w}{\partial y}\, d\boldsymbol{x} + \int_\Omega \frac{\partial v}{\partial y} w\, d\boldsymbol{x} = \int_{\partial\Omega} vwn_2\, d\boldsymbol{\sigma}. \qquad (\star_2)$$

Notation: $n = 2$, $v : \overline{\Omega} \subset \mathbb{R}^2 \to \mathbb{R}$ a smooth function.

- $\nabla v := \left(\frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}\right)^\top$ = gradient of v
- $\Delta v := \nabla \cdot \nabla v := \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}$ = Laplacian of v
- $\frac{\partial v}{\partial \boldsymbol{n}} := \nabla v \cdot \boldsymbol{n} = \frac{\partial v}{\partial x} n_1 + \frac{\partial v}{\partial y} n_2$ = normal derivative of v

Integration by parts (Green's formula)

Assume that $\boldsymbol{v} = (v_1, v_2)^\top$, $v_1, v_2 : \overline{\Omega} \subset \mathbb{R}^2 \to \mathbb{R}$ and $p : \overline{\Omega} \subset \mathbb{R}^2 \to \mathbb{R}$ are all smooth functions. Then we have

$$\int_\Omega \boldsymbol{v} \cdot \nabla p \, d\boldsymbol{x} = \int_{\partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n}) p \, d\boldsymbol{\sigma} - \int_\Omega (\nabla \cdot \boldsymbol{v}) p \, d\boldsymbol{x}$$

Proof. By using $(\star_1)$ and $(\star_2)$, we have

$$\begin{aligned}
\int_\Omega \boldsymbol{v} \cdot \nabla p \, d\boldsymbol{x} &= \int_\Omega \left(v_1 \frac{\partial p}{\partial x} + v_2 \frac{\partial p}{\partial y} \right) d\boldsymbol{x} \\
&= \int_{\partial\Omega} v_1 p n_1 \, d\boldsymbol{\sigma} - \int_\Omega \frac{\partial v_1}{\partial x} p \, d\boldsymbol{x} + \int_{\partial\Omega} v_2 p n_2 \, d\boldsymbol{\sigma} - \int_\Omega \frac{\partial v_2}{\partial y} p \, d\boldsymbol{x} \\
&= \int_{\partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n}) p \, d\boldsymbol{\sigma} - \int_\Omega (\nabla \cdot \boldsymbol{v}) p \, d\boldsymbol{x}.
\end{aligned}$$

The total derivative of L

Let us go back to the total derivative,

$$\frac{dL}{d\varepsilon} = \frac{\partial L}{\partial v}\eta + \frac{\partial L}{\partial v_x}\eta_x + \frac{\partial L}{\partial v_y}\eta_y = \frac{\partial L}{\partial v}\eta + \left(\frac{\partial L}{\partial v_x}, \frac{\partial L}{\partial v_y}\right)^\top \cdot \nabla\eta$$

By the integration by parts (Green's formula), we obtain

$$\begin{aligned} 0 &= \int_\Omega \frac{dL}{d\varepsilon}\Big|_{\varepsilon=0} d\boldsymbol{x} = \int_\Omega \frac{\partial L}{\partial u}\eta + \big(\frac{\partial L}{\partial u_x}, \frac{\partial L}{\partial u_y}\big)^\top \cdot \nabla\eta \, d\boldsymbol{x} \\ &= \int_\Omega \frac{\partial L}{\partial u}\eta \, d\boldsymbol{x} + \int_{\partial\Omega} \Big(\big(\frac{\partial L}{\partial u_x}, \frac{\partial L}{\partial u_y}\big)^\top \cdot \boldsymbol{n}\Big)\eta \, d\boldsymbol{\sigma} \\ &\quad - \int_\Omega \Big(\nabla \cdot \big(\frac{\partial L}{\partial u_x}, \frac{\partial L}{\partial u_y}\big)^\top\Big)\eta \, d\boldsymbol{x}, \qquad (\star) \end{aligned}$$

where $L(x, y, v, v_x, v_y) \rightsquigarrow L(x, y, u, u_x, u_y)$ when $\varepsilon = 0$. Taking arbitrary smooth functions η's with $\eta(\boldsymbol{x}) = 0$ on $\partial\Omega$, we have

$$\int_\Omega \eta \left\{ \frac{\partial L}{\partial u} - \nabla \cdot \big(\frac{\partial L}{\partial u_x}, \frac{\partial L}{\partial u_y}\big)^\top \right\} d\boldsymbol{x} = 0.$$

The Euler-Lagrange equation

- *According to the fundamental lemma of calculus of variations (see next page), we obtain the functional derivative of E at u,*

$$\frac{\partial L}{\partial u} - \nabla \cdot \left(\frac{\partial L}{\partial u_x}, \frac{\partial L}{\partial u_y}\right)^\top = 0 \quad \text{in } \Omega, \qquad (\star\star)$$

  which is the so-called Euler-Lagrange equation.

- By substituting $(\star\star)$ into $(\star)$, we have

$$\int_{\partial\Omega} \eta \left(\frac{\partial L}{\partial u_x} n_1 + \frac{\partial L}{\partial u_y} n_2 \right) d\sigma = 0,$$

  for any smooth function η on $\overline{\Omega}$, which implies the homogeneous Neumann boundary condition (BC),

$$\frac{\partial L}{\partial u_x} n_1 + \frac{\partial L}{\partial u_y} n_2 = 0 \quad \text{on } \partial\Omega.$$

Fundamental lemma of calculus of variations

If F is a continuous real-valued function on an open bounded set $\Omega \subset \mathbb{R}^2$ and satisfies

$$\int_\Omega F(\boldsymbol{x})G(\boldsymbol{x})\,d\boldsymbol{x} = 0$$

for all compactly supported smooth functions G on Ω, then we have

$$F(\boldsymbol{x}) = 0 \quad \textit{for all } \boldsymbol{x} \in \Omega.$$

Note: *If F is continuous on the closure $\overline{\Omega}$, then we require only that G vanishes on the boundary $\partial\Omega$ of Ω to ensure the assertion in the above lemma.*

Euler-Lagrange equation of the ROF model

Consider the regularized minimization problem:

$$\min_{u\in\mathcal{V}}\Big(F(u)+\frac{\lambda}{2}\int_\Omega \big(u(\boldsymbol{x})-f(\boldsymbol{x})\big)^2\,d\boldsymbol{x}\Big),$$

where $\mathcal{V}$ is a suitable space and $\lambda > 0$ is the regularization parameter.

ROF regularizer: $F(u)=\int_\Omega |\nabla u|\,d\boldsymbol{x}=\int_\Omega \sqrt{u_x^2+u_y^2}\,d\boldsymbol{x}$, we have

$$L(x,y,u,u_x,u_y)=L(u,u_x,u_y)=\sqrt{u_x^2+u_y^2}+\frac{\lambda}{2}(u-f)^2,$$

which leads to the Euler-Lagrange equation with the Neumann BC,

$$-\nabla\cdot\Big(\frac{\nabla u}{|\nabla u|}\Big)+\lambda u=\lambda f \quad \text{in } \Omega, \qquad \frac{\partial u}{\partial \boldsymbol{n}}=0 \quad \text{on } \partial\Omega.$$

The homogeneous Neumann BC comes from

$$0=\frac{\partial L}{\partial u_x}n_1+\frac{\partial L}{\partial u_y}n_2=\Big(\frac{\partial L}{\partial u_x},\frac{\partial L}{\partial u_y}\Big)\cdot\boldsymbol{n}=\frac{\nabla u}{|\nabla u|}\cdot\boldsymbol{n}=\frac{1}{|\nabla u|}\frac{\partial u}{\partial \boldsymbol{n}} \quad \text{on } \partial\Omega.$$

(If $|\nabla u|=0 \Rightarrow \nabla u=\mathbf{0} \Rightarrow \nabla u\cdot\boldsymbol{n}=0$; otherwise $\frac{\partial u}{\partial \boldsymbol{n}}=0$)

The Euler-Lagrange equations of a Tikhonov regularizer

Consider the regularized minimization problem:

$$\min_{u \in \mathcal{V}} \left(F(u) + \frac{\lambda}{2} \int_\Omega \big(u(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2 \, d\boldsymbol{x} \right),$$

where $\mathcal{V}$ is a suitable space and $\lambda > 0$ is the regularization parameter.

Tikhonov quadratic regularizer: $F(u) = \frac{1}{2} \int_\Omega u^2 \, d\boldsymbol{x}$, we have

$$L(x, y, u, u_x, u_y) = L(u, u_x, u_y) = \frac{1}{2} u^2 + \frac{\lambda}{2} (u - f)^2,$$

which implies the Euler-Lagrange equation,

$$u + \lambda u = \lambda f \quad \text{in } \Omega \implies u = \frac{\lambda}{1 + \lambda} f \quad \text{in } \Omega,$$

but without any boundary condition because $\dfrac{\partial L}{\partial u_x} = 0 = \dfrac{\partial L}{\partial u_y}$.

Another Tikhonov quadratic regularizer

Consider the regularized minimization problem:

$$\min_{u\in\mathcal{V}}\left(F(u)+\frac{\lambda}{2}\int_\Omega \big(u(\boldsymbol{x})-f(\boldsymbol{x})\big)^2\,d\boldsymbol{x}\right),$$

where $\mathcal{V}$ is a suitable space and $\lambda > 0$ is the regularization parameter.

Tikhonov regularizer: Let $F(u)=\frac{1}{2}\int_\Omega |\nabla u|^2\,d\boldsymbol{x}=\frac{1}{2}\int_\Omega u_x^2+u_y^2\,d\boldsymbol{x}$.
Then we have

$$L(x,y,u,u_x,u_y)=L(u,u_x,u_y)=\frac{1}{2}(u_x^2+u_y^2)+\frac{\lambda}{2}(u-f)^2,$$

which implies the Euler-Lagrange equation,

$$-\nabla\cdot\nabla u+\lambda u=\lambda f \quad \text{in } \Omega \implies -\frac{1}{\lambda}\Delta u+u=f \quad \text{in } \Omega,$$

with the homogeneous Neumann BC,

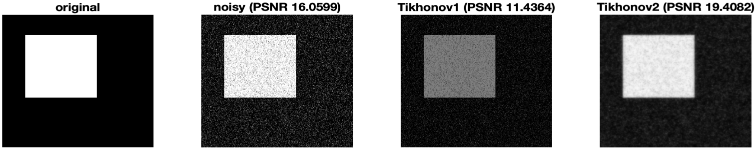
$$0=\frac{\partial L}{\partial u_x}n_1+\frac{\partial L}{\partial u_y}n_2=\nabla u\cdot\boldsymbol{n}=\frac{\partial u}{\partial \boldsymbol{n}} \quad \text{on } \partial\Omega.$$

Numerical results of the two Tikhonov models

Consider the two Tikhonov models:

$$\min_{u\in\mathcal{V}}\left(F(u)+\frac{\lambda}{2}\int_\Omega (u(\boldsymbol{x})-f(\boldsymbol{x}))^2\,d\boldsymbol{x}\right),\ F(u)=\frac{1}{2}\int_\Omega u^2\,d\boldsymbol{x}\ \&\ \frac{1}{2}\int_\Omega |\nabla u|^2\,d\boldsymbol{x}.$$

- In the first model, there is no regularization of any kind, since $u(\boldsymbol{x})=(\frac{\lambda}{1+\lambda})f(\boldsymbol{x})$ in Ω. Obviously, this is a wrong choice.
- *In the second model, the function space is $\mathcal{V}:=H^1(\Omega)$. However, there is too much regularization. In fact, the image u belongs to $H^1(\Omega)$, which cannot present discontinuities such as edges or boundaries of objects.*



$\lambda=1,\quad$ (T1): $F(u)=\frac{1}{2}\int_\Omega u^2\,d\boldsymbol{x},\quad$ (T2): $F(u)=\frac{1}{2}\int_\Omega |\nabla u|^2\,d\boldsymbol{x}$

Model Tikhonov2 for different λ's

$$\lambda = 1, 1/2, 1/3, 1/4; F(u) = \frac{1}{2} \int_\Omega |\nabla u|^2 \, dx$$

Three indices to measure the quality

Below are three indices to measure the produced image quality and evaluate denoising performance. *Let $\tilde{u}$ be the clean digital grayscale image of pixel size $M \times N$, $\overline{u}$ be the mean intensity of the clean image, and u be the produced (denoised) image.*

$$
\begin{aligned}
MSE(\tilde{u}, u) &:= \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (\tilde{u}_{i,j} - u_{i,j})^2 && \text{(mean squared error)} \\
PSNR &:= 10 \log_{10} \left(\frac{255^2}{MSE(\tilde{u}, u)} \right) && \text{(peak signal to noise ratio)} \\
SNR &:= 10 \log_{10} \left(\frac{MSE(\tilde{u}, \overline{u})}{MSE(\tilde{u}, u)} \right) && \text{(signal to noise ratio)}
\end{aligned}
$$

In general, the higher the value of $PSNR$ the better the quality of the produced image.

There is another index, structural similarity ($SSIM$). The maximum value of $SSIM$ is 1.

Nonlinear PDE-based denoising algorithm

The boundary value problem (BVP) of the ROF model is given by

$$\begin{aligned} -\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) + \lambda u &= \lambda f \quad \text{in } \Omega, \\ \frac{\partial u}{\partial \boldsymbol{n}} &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

Since the energy functional of ROF model is convex, the solution of the BVP is the minimizer of the ROF minimization model.

The numerical solution of the above BVP can be obtained by evolving a finite difference approximation of the parabolic PDE with the homogeneous Neumann BC to reach a steady state:

$$\begin{cases} \overbrace{\frac{\partial u}{\partial t} - \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) + \lambda u = \lambda f \quad \text{for } (t, \boldsymbol{x}) \in (0, T) \times \Omega,}^{Heat-type\ equation} \\ \underbrace{\nabla u \cdot \boldsymbol{n} = 0 \quad \text{for } t \in [0, T] \text{ and } \boldsymbol{x} \in \partial\Omega}_{homogeneous\ Neumann\ BC} \oplus \underbrace{u(0, \boldsymbol{x}) = f(\boldsymbol{x}) \text{ for } \boldsymbol{x} \in \overline{\Omega}.}_{initial\ condition} \end{cases}$$

Numerical differentiation: 1-D

Let $v : [a, b] \to \mathbb{R}$ and let $a = x_0 < x_1 < \cdots < x_N = b$ be a uniform partition of $[a, b]$ with grid size $h = (b - a)/N > 0$.

Forward difference for $v'(x_i)$: Assume that $v \in C^2[a, b]$. Then for $i = 1, 2, \cdots, N-1$, by Taylor's theorem, we have

$v(x_i + h) = v(x_i) + v'(x_i)h + \frac{1}{2}v''(\xi_i)h^2$ for some $\xi_i \in (x_i, x_i + h)$.

$\therefore v'(x_i) = \frac{1}{h}\big(v(x_i + h) - v(x_i)\big) - \frac{1}{2}v''(\xi_i)h$

$\therefore v'(x_i) \approx \frac{1}{h}\big(v(x_{i+1}) - v(x_i)\big)$, it is a first-order approximation!

Backward difference for $v'(x_i)$: Assume that $v \in C^2[a, b]$. Then for $i = 1, 2, \cdots, N-1$, by Taylor's theorem, we have

$v(x_i - h) = v(x_i) - v'(x_i)h + \frac{1}{2}v''(\xi_i)h^2$ for some $\xi_i \in (x_i - h, x_i)$.

$\therefore v'(x_i) = \frac{1}{h}\big(v(x_i) - v(x_i - h)\big) + \frac{1}{2}v''(\xi_i)h$

$\therefore v'(x_i) \approx \frac{1}{h}\big(v(x_i) - v(x_{i-1})\big)$, it is a first-order approximation!

Numerical differentiation (cont'd)

Central difference for $v'(x_i)$: Assume that $v \in C^3[a,b]$. Then for $i = 1, 2, \cdots, N-1$, by Taylor's theorem, we have

$$v(x_i + h) = v(x_i) + v'(x_i)h + \frac{1}{2}v''(x_i)h^2 + \frac{1}{6}v^{(3)}(\xi_{i1})h^3,$$

$$v(x_i - h) = v(x_i) - v'(x_i)h + \frac{1}{2}v''(x_i)h^2 - \frac{1}{6}v^{(3)}(\xi_{i2})h^3,$$

for some $\xi_{i1} \in (x_i, x_i + h)$ and $\xi_{i2} \in (x_i - h, x_i)$.

Subtracting the second equation from the first equation, we have

$$v(x_i + h) - v(x_i - h) = 2v'(x_i)h + \frac{1}{6}h^3\left(v^{(3)}(\xi_{i1}) + v^{(3)}(\xi_{i2})\right).$$

$$\therefore\ v'(x_i) = \frac{1}{2h}\left(v(x_i + h) - v(x_i - h)\right) - \frac{1}{6}h^2\frac{1}{2}\left(v^{(3)}(\xi_{i1}) + v^{(3)}(\xi_{i2})\right)$$

$\because\ \frac{1}{2}\left(v^{(3)}(\xi_{i1}) + v^{(3)}(\xi_{i2})\right)$ is between $v^{(3)}(\xi_{i1})$ and $v^{(3)}(\xi_{i2})$

$\therefore$ By the intermediate value theorem, $\exists\ \xi_i \in (x_i - h, x_i + h)$ such that

$$v^{(3)}(\xi_i) = \frac{1}{2}\left(v^{(3)}(\xi_{i1}) + v^{(3)}(\xi_{i2})\right)$$

$$\therefore\ v'(x_i) = \frac{1}{2h}\left(v(x_i + h) - v(x_i - h)\right) - \frac{1}{6}h^2 v^{(3)}(\xi_i)$$

$\therefore\ v'(x_i) \approx \frac{1}{2h}\left(v(x_{i+1}) - v(x_{i-1})\right)$, 2nd-order approximation!

Numerical differentiation (cont'd)

Central difference for $v''(x_i)$: Assume that $v \in C^4[a,b]$. Then for $i = 1, 2, \cdots, N-1$, by Taylor's theorem, we have

$$v(x_i + h) = v(x_i) + v'(x_i)h + \frac{1}{2}v''(x_i)h^2 + \frac{1}{6}v^{(3)}(x_i)h^3 + \frac{1}{24}v^{(4)}(\xi_{i1})h^4,$$

$$v(x_i - h) = v(x_i) - v'(x_i)h + \frac{1}{2}v''(x_i)h^2 - \frac{1}{6}v^{(3)}(x_i)h^3 + \frac{1}{24}v^{(4)}(\xi_{i2})h^4,$$

for some $\xi_{i1} \in (x_i, x_i + h)$ and $\xi_{i2} \in (x_i - h, x_i)$.

Adding these two equations, we have

$$v(x_i + h) + v(x_i - h) = 2v(x_i) + v''(x_i)h^2 + \frac{1}{24}\{v^{(4)}(\xi_{i1}) + v^{(4)}(\xi_{i2})\}h^4.$$

$\therefore$

$$v''(x_i) = \frac{1}{h^2}\{v(x_i + h) - 2v(x_i) + v(x_i - h)\} - \frac{h^2}{24}\{v^{(4)}(\xi_{i1}) + v^{(4)}(\xi_{i2})\}$$

$\because v \in C^4[a,b]$, $\frac{1}{2}\{v^{(4)}(\xi_{i1}) + v^{(4)}(\xi_{i2})\}$ between $v^{(4)}(\xi_{i1})$ & $v^{(4)}(\xi_{i2})$

$\therefore$ By IVT, $\exists\ \xi_i$ between ξ_{i1} and ξ_{i2} ($\Rightarrow \xi_i \in (x_i - h, x_i + h)$) such that

$$v^{(4)}(\xi_i) = \frac{1}{2}\{v^{(4)}(\xi_{i1}) + v^{(4)}(\xi_{i2})\}$$

$$\therefore v''(x_i) = \frac{1}{h^2}\{v(x_i + h) - 2v(x_i) + v(x_i - h)\} - \frac{1}{12}h^2 v^{(4)}(\xi_i)$$

$\therefore v''(x_i) \approx \frac{1}{h^2}\{v(x_{i+1}) - 2v(x_i) + v(x_{i-1})\}$, 2nd-order approximation!

Let $u_{i,j}^n$ denote an approximation to $u(t_n, x_i, y_j)$

- $\dfrac{\partial u}{\partial x}(t_n, x_i, y_j) \approx \nabla_x^+ u_{i,j}^n := \dfrac{u_{i+1,j}^n - u_{i,j}^n}{h}$ (forward difference in x)
- $\dfrac{\partial u}{\partial x}(t_n, x_i, y_j) \approx \nabla_x^- u_{i,j}^n := \dfrac{u_{i,j}^n - u_{i-1,j}^n}{h}$ (backward difference in x)
- $\dfrac{\partial u}{\partial x}(t_n, x_i, y_j) \approx \nabla_x u_{i,j}^n := \dfrac{u_{i+1,j}^n - u_{i-1,j}^n}{2h} = \dfrac{1}{2}\left(\nabla_x^+ u_{i,j}^n + \nabla_x^- u_{i,j}^n\right)$
  (central difference in x)
- $\dfrac{\partial u}{\partial y}(t_n, x_i, y_j) \approx \nabla_y^+ u_{i,j}^n := \dfrac{u_{i,j+1}^n - u_{i,j}^n}{h}$ (forward difference in y)
- $\dfrac{\partial u}{\partial y}(t_n, x_i, y_j) \approx \nabla_y^- u_{i,j}^n := \dfrac{u_{i,j}^n - u_{i,j-1}^n}{h}$ (backward difference in y)
- $\dfrac{\partial u}{\partial y}(t_n, x_i, y_j) \approx \nabla_y u_{i,j}^n := \dfrac{u_{i,j+1}^n - u_{i,j-1}^n}{2h} = \dfrac{1}{2}\left(\nabla_y^+ u_{i,j}^n + \nabla_y^- u_{i,j}^n\right)$
  (central difference in y)

Central differences for second derivative

- Central difference for second derivative in x:

$$
\begin{aligned}
\nabla_x^-(\nabla_x^+ u_{i,j}^n) &= \nabla_x^-\left(\frac{u_{i+1,j}^n - u_{i,j}^n}{h}\right) = \frac{1}{h}\left(\nabla_x^- u_{i+1,j}^n - \nabla_x^- u_{i,j}^n\right) \\
&= \frac{1}{h}\left(\frac{u_{i+1,j}^n - u_{i,j}^n}{h} - \frac{u_{i,j}^n - u_{i-1,j}^n}{h}\right) \\
&= \frac{1}{h^2}\left(u_{i+1,j}^n - 2u_{i,j}^n + u_{i-1,j}^n\right) \approx \frac{\partial^2 u}{\partial x^2}(t_n, x_i, y_j).
\end{aligned}
$$

- Central difference for second derivative in y:

$$
\nabla_y^-(\nabla_y^+ u_{i,j}^n) = \frac{1}{h^2}\left(u_{i,j+1}^n - 2u_{i,j}^n + u_{i,j-1}^n\right) \approx \frac{\partial^2 u}{\partial y^2}(t_n, x_i, y_j).
$$

- $\nabla_x^+(\nabla_x^- u_{i,j}^n) = \nabla_x^-(\nabla_x^+ u_{i,j}^n)$, will also be denoted as $\nabla_x^2 u_{i,j}^n$.
  $\nabla_y^+(\nabla_y^- u_{i,j}^n) = \nabla_y^-(\nabla_y^+ u_{i,j}^n)$, will also be denoted as $\nabla_y^2 u_{i,j}^n$.

Forward Euler in time t

We will consider a finite difference scheme for approximating the solution of the IBVP for the Euler-Lagrange equation:

$$\frac{\partial u}{\partial t} - \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) + \lambda u = \lambda f \quad \text{for } (t, \boldsymbol{x}) \in (0, T) \times \Omega,$$

$$u(0, \boldsymbol{x}) = f(\boldsymbol{x}) \text{ for } \boldsymbol{x} \in \overline{\Omega},$$

$$\nabla u \cdot \boldsymbol{n} = 0 \quad \text{for } t \in [0, T] \text{ and } \boldsymbol{x} \in \partial\Omega.$$

Suppose that the image domain is given by $\overline{\Omega} = [0, 1] \times [0, 1]$. Let $x_i = ih$ and $y_j = jh$, $i, j = 0, 1, \cdots, N$, with $h = 1/N$, and $t_n = n\Delta t$. Let $f_{i,j} := f(x_i, y_j)$ and $u_{i,j}^n$ be the difference approximation to $u(t_n, x_i, y_j)$.

Forward Euler in time t:

$$\begin{aligned} \frac{\partial u}{\partial t}(t_n, x_i, y_j) &= \frac{1}{\Delta t}\left(u(t_{n+1}, x_i, y_j) - u(t_n, x_i, y_j)\right) - \frac{1}{2}\frac{\partial^2 u}{\partial t^2}(\tau_i, x_i, y_j)\Delta t \\ &\approx \frac{1}{\Delta t}\left(u_{i,j}^{n+1} - u_{i,j}^n\right). \end{aligned}$$

The forward Euler finite difference scheme

The proposed explicit finite difference scheme is given by:

$$\frac{u_{i,j}^{n+1} - u_{i,j}^n}{\Delta t} = \lambda(f_{i,j} - u_{i,j}^n) + \nabla_x^- \left(\frac{\nabla_x^+ u_{i,j}^n}{\sqrt{(\nabla_x^+ u_{i,j}^n)^2 + (m(\nabla_y^+ u_{i,j}^n, \nabla_y^- u_{i,j}^n))^2}} \right)$$
$$+ \nabla_y^- \left(\frac{\nabla_y^+ u_{i,j}^n}{\sqrt{(\nabla_y^+ u_{i,j}^n)^2 + (m(\nabla_x^+ u_{i,j}^n, \nabla_x^- u_{i,j}^n))^2}} \right), \quad 1 \le i,j \le N-1,$$

$$u_{0,j}^n = u_{1,j}^n,\ u_{N,j}^n = u_{N-1,j}^n,\ u_{i,0}^n = u_{i,1}^n,\ u_{i,N}^n = u_{i,N-1}^n, \quad 0 \le i,j \le N.$$

where $m(a,b) = \left(\frac{\text{sign}\, a + \text{sign}\, b}{2}\right) \min\{|a|,|b|\}$ *is the minmod operator;* see [ROF 1992] for more details.

- *The forward Euler scheme is conditionally stable, we need* $\Delta t/h^2 \le c$.
- Numerous other algorithms have been proposed to solve the *TV* denoising minimization problem, e.g., *the split Bregman iterations.*

Rescaling the finite difference scheme

Let $\delta_x^+ u_{i,j}^n := u_{i+1,j}^n - u_{i,j}^n$, $\delta_x^- u_{i,j}^n := u_{i,j}^n - u_{i-1,j}^n$, $\delta_y^+ u_{i,j}^n := u_{i,j+1}^n - u_{i,j}^n$, $\delta_y^- u_{i,j}^n := u_{i,j}^n - u_{i,j-1}^n$. Then the proposed finite difference scheme can be rewritten as

$$\frac{u_{i,j}^{n+1} - u_{i,j}^n}{\Delta t} = \lambda(f_{i,j} - u_{i,j}^n) + \frac{1}{h}\delta_x^- \left(\frac{\delta_x^+ u_{i,j}^n}{\sqrt{(\delta_x^+ u_{i,j}^n)^2 + (m(\delta_y^+ u_{i,j}^n, \delta_y^- u_{i,j}^n))^2}} \right)$$
$$+ \frac{1}{h}\delta_y^- \left(\frac{\delta_y^+ u_{i,j}^n}{\sqrt{(\delta_y^+ u_{i,j}^n)^2 + (m(\delta_x^+ u_{i,j}^n, \delta_x^- u_{i,j}^n))^2}} \right), \quad 1 \le i,j \le N-1,$$

$$u_{0,j}^n = u_{1,j}^n,\ u_{N,j}^n = u_{N-1,j}^n,\ u_{i,0}^n = u_{i,1}^n,\ u_{i,N}^n = u_{i,N-1}^n, \quad 0 \le i,j \le N.$$

Let $A_{i,j}^n := \dfrac{\delta_x^+ u_{i,j}^n}{\sqrt{(\delta_x^+ u_{i,j}^n)^2 + (m(\delta_y^+ u_{i,j}^n, \delta_y^- u_{i,j}^n))^2}}$ and

$$B_{i,j}^n := \frac{\delta_y^+ u_{i,j}^n}{\sqrt{(\delta_y^+ u_{i,j}^n)^2 + (m(\delta_x^+ u_{i,j}^n, \delta_x^- u_{i,j}^n))^2}}.$$

Rescaling the finite difference scheme (cont'd)

Then we have

$$\frac{u_{i,j}^{n+1} - u_{i,j}^{n}}{\Delta t} = \lambda(f_{i,j} - u_{i,j}^{n}) + \tfrac{1}{h}\delta_x^- A_{i,j}^n + \tfrac{1}{h}\delta_y^- B_{i,j}^n, \quad 1 \le i,j \le N-1,$$

$$u_{0,j}^n = u_{1,j}^n,\ u_{N,j}^n = u_{N-1,j}^n,\ u_{i,0}^n = u_{i,1}^n,\ u_{i,N}^n = u_{i,N-1}^n, \quad 0 \le i,j \le N.$$

Setting $\widetilde{\Delta t} = \dfrac{\Delta t}{h}$ and $\widetilde{\lambda} = h\lambda$, the first equation becomes

$$\frac{u_{i,j}^{n+1} - u_{i,j}^{n}}{\widetilde{\Delta t}} = \widetilde{\lambda}(f_{i,j} - u_{i,j}^{n}) + \delta_x^- A_{i,j}^n + \delta_y^- B_{i,j}^n, \quad 1 \le i,j \le N-1.$$

Rearranging the equation, we finally obtain

$$u_{i,j}^{n+1} = u_{i,j}^{n} + \widetilde{\Delta t}\widetilde{\lambda}(f_{i,j} - u_{i,j}^{n}) + \widetilde{\Delta t}\delta_x^- A_{i,j}^n + \widetilde{\Delta t}\delta_y^- B_{i,j}^n, \quad 1 \le i,j \le N-1.$$

A uniform partition of $\Omega = (0,1) \times (0,1)$

Let " $\bullet$ " denote an arbitrary point (x, y) in $\overline{\Omega}$.

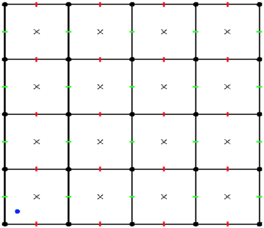
(1) In usual finite differences, the grid points (x_i, y_j) locate at " $\bullet$ ".

(2) *In image processing, however, a digital image is usually stored as a matrix. Thus, it is more convenient to use the "cell-centered grids," i.e., grid points (x_i, y_j) located at " $\times$ " with the coordinates*

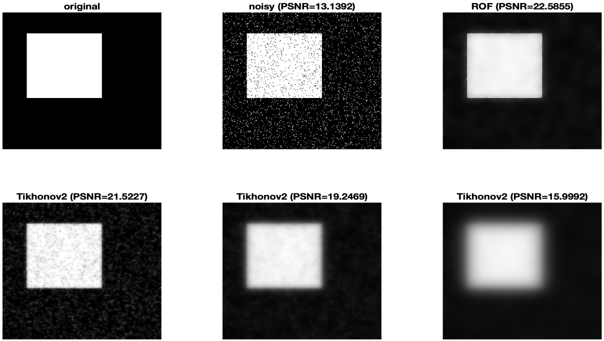
$$x_i = \frac{h}{2} + (i-1)h, \;\; y_j = \frac{h}{2} + (j-1)h, \;\; i,j = (0), 1, \cdots N, (N+1).$$

And the homogeneous Neumann BC implies

$$u^n_{0,j} = u^n_{1,j},\; u^n_{N+1,j} = u^n_{N,j},\; u^n_{i,0} = u^n_{i,1},\; u^n_{i,N+1} = u^n_{i,N}, \;\; 1 \leq i,j \leq N.$$

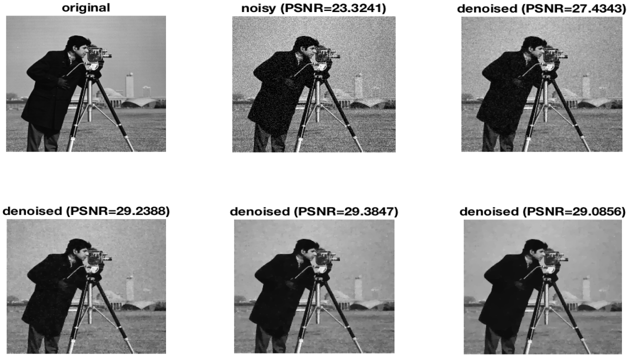


ROF model versus Tikhonov2 model



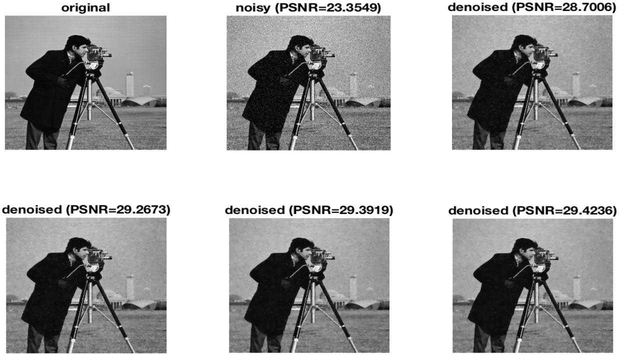
ROF: $\lambda = 1, 100$*th step;* *Tikhonov2:* $\lambda = 1, 0.1, 0.01$

ROF finite difference solutions at different steps (cameraman)



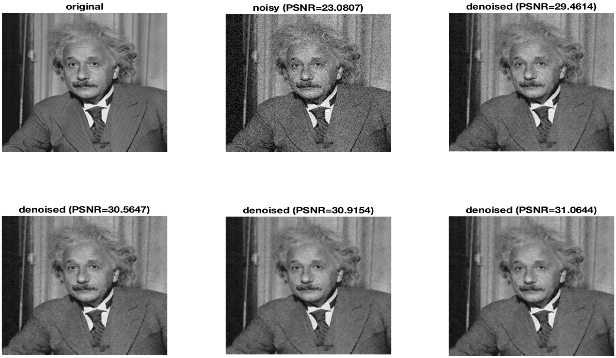
Gaussian noise $(0, 0.005)$, $h = 1/256$, $\widetilde{\lambda} = h\lambda = 0.05$, $\widetilde{\Delta t} = \Delta t/h = 0.01$, *at* $500, 1000, 1500, 2000$*-th steps*

ROF finite difference solutions of different λ's (cameraman)



Gaussian noise $(0, 0.005)$, $h = 1/256$, $\widetilde{\lambda} = h\lambda = 1/10, 1/20, 1/30, 1/40$,
$\widetilde{\Delta t} = \Delta t/h = 0.01$, *at* 1000-*th step*

ROF finite difference solutions of different λ's (Einstein)



Gaussian noise $(0, 0.005)$, $h = 1/340$, $\widetilde{\lambda} = h\lambda = 1/10, 1/20, 1/30, 1/40$,
$\widetilde{\Delta t} = \Delta t/h = 0.01$, *at* 1000-*th step*

An alternating direction approach: split Bregman method

Below, we introduce the split Bregman method which is an alternating direction approach to solve the ROF model. First, using the cell-centered grids of $\overline{\Omega}$, we approximate the total variation term by the Riemann sum:

$$\|u\|_{TV(\Omega)} = \int_\Omega |\nabla u| \, d\boldsymbol{x} := \int_\Omega \sqrt{u_x^2 + u_y^2} \, d\boldsymbol{x} \approx h^2 \sum_{i=1}^{N} \sum_{j=1}^{N} |\nabla_h u_{i,j}|.$$

Here we define the discrete gradient operator ∇_h by

$$\nabla_h u_{i,j} := [\nabla_x u_{i,j}, \nabla_y u_{i,j}]^\top$$

and recall that

$$\nabla_x u_{i,j} = \frac{u_{i+1,j} - u_{i-1,j}}{2h}, \qquad \nabla_y u_{i,j} = \frac{u_{i,j+1} - u_{i,j-1}}{2h}, \quad 1 \le i,j \le N,$$

$$u_{0,j} = u_{1,j},\ u_{N+1,j} = u_{N,j},\ u_{i,0} = u_{i,1},\ u_{i,N+1} = u_{i,N}, \quad 1 \le i,j \le N.$$

The constrained minimization of the ROF model

Introducing the new unknown vector function $\boldsymbol{d}(\boldsymbol{x}) = \nabla u(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$, we have the constrained minimization problem:

$$\min_{u,\boldsymbol{d}} \left(\int_\Omega |\boldsymbol{d}(\boldsymbol{x})| \, d\boldsymbol{x} + \frac{\lambda}{2} \int_\Omega \left(u(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2 d\boldsymbol{x} \right) \text{ s.t. } \boldsymbol{d}(\boldsymbol{x}) = \nabla u(\boldsymbol{x}),$$

where $|\cdot| := \|\cdot\|_2$ in $\mathbb{R}^2$. Therefore, the approximate constrained minimization of the ROF model can be posed as follows:

$$\min_{u,\boldsymbol{d}} \left(\hbar^2 \sum_{i,j=1}^{N} |\boldsymbol{d}_{i,j}| + \hbar^2 \frac{\lambda}{2} \sum_{i,j=1}^{N} (f_{i,j} - u_{i,j})^2 \right) \text{ s.t. } \boldsymbol{d}_{i,j} = \nabla_h u_{i,j} = \begin{bmatrix} \nabla_x u_{i,j} \\ \nabla_y u_{i,j} \end{bmatrix},$$

where u and $\boldsymbol{d}$ denote the set of all $u_{i,j}$ and $d_{i,j}$. *Introducing a penalty parameter $\gamma > 0$, we obtain the unconstrained problem:*

$$\min_{u,\boldsymbol{d},\boldsymbol{b}} \left(\sum_{i,j=1}^{N} |\boldsymbol{d}_{i,j}| + \frac{\lambda}{2} \sum_{i,j=1}^{N} (f_{i,j} - u_{i,j})^2 + \frac{\gamma}{2} \sum_{i,j=1}^{N} |\boldsymbol{d}_{i,j} - \nabla_h u_{i,j} - \boldsymbol{b}_{i,j}|^2 \right),$$

where $\boldsymbol{b}$ (denotes the set of all $\boldsymbol{b}_{i,j}$) is an auxiliary variable, which can be expressed in terms of u and $\boldsymbol{d}$, related to the Bregman iterations.

An alternating direction approach: split Bregman method

Goldstein and Osher (2009) proposed to solve the above problem by an alternating direction approach; see also Getreuer (2012):

u**-subproblem:** With $\boldsymbol{d}$ and $\boldsymbol{b}$ fixed, we solve

$$u^{k+1} = \arg\min_u \left(\frac{\lambda}{2} \sum_{i,j} (f_{i,j} - u_{i,j})^2 + \frac{\gamma}{2} \sum_{i,j} |\boldsymbol{d}^k_{i,j} - \nabla_h u_{i,j} - \boldsymbol{b}^k_{i,j}|^2 \right),$$

where the superscript k denotes the values evaluated at k-iteration. It can be viewed as the approximation of the minimization problem:

$$\min_u \frac{\lambda}{2} \int_\Omega (f-u)^2 \, d\boldsymbol{x} + \frac{\gamma}{2} \int_\Omega |\boldsymbol{d}^k - \nabla u - \boldsymbol{b}^k|^2 \, d\boldsymbol{x}.$$

The associated Euler-Lagrange equation of the above minimization problem (also called the screened Poisson equation) is given by

$$\lambda u - \gamma \nabla \cdot \nabla u = \lambda f - \gamma \nabla \cdot (\boldsymbol{d}^k - \boldsymbol{b}^k),$$

where ∇u is the gradient of u, $\nabla \cdot \boldsymbol{v}$ is the divergence of vector function $\boldsymbol{v}$, and $\Delta u := \nabla^2 u := \nabla \cdot \nabla u$ is the Laplacian of u.

The discrete screened Poisson equation

The discrete screened Poisson equation is given by

$$\lambda u_{i,j} - \gamma \nabla_h^2 u_{i,j} = \lambda f_{i,j} - \gamma \nabla_h \cdot (\boldsymbol{d}_{i,j}^k - \boldsymbol{b}_{i,j}^k), \quad 1 \le i,j \le N,$$

which should be supplemented with the homogeneous Neumann BC:

$$u_{0,j} = u_{1,j}, u_{N+1,j} = u_{N,j}, u_{i,0} = u_{i,1}, u_{i,N+1} = u_{i,N}, \quad 1 \le i,j \le N.$$

- The term $\Delta_h u_{i,j} := \nabla_h^2 u_{i,j} := \nabla_h^- \cdot \nabla_h^+ u_{i,j}$

$$\begin{aligned} \nabla_h^- \cdot \nabla_h^+ u_{i,j} &= (\nabla_x^-, \nabla_y^-)^\top \cdot (\nabla_x^+ u_{i,j}, \nabla_y^+ u_{i,j})^\top = \nabla_x^- \nabla_x^+ u_{i,j} + \nabla_y^- \nabla_y^+ u_{i,j} \\ &= \frac{1}{h^2}\Big(\big(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}\big) + \big(u_{i,j+1} - 2u_{i,j} + u_{i,j-1}\big) \Big) \\ &= \frac{1}{h^2}\Big(-4u_{i,j} + u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1} \Big). \end{aligned}$$

- Let $\boldsymbol{g}_{i,j}^k = (g_{1,i,j}^k, g_{2,i,j}^k)^\top := \boldsymbol{d}_{i,j}^k - \boldsymbol{b}_{i,j}^k$. Then

$$\nabla_h \cdot \boldsymbol{g}_{i,j}^k = \nabla_x g_{1,i,j}^k + \nabla_y g_{2,i,j}^k = \frac{g_{1,i+1,j}^k - g_{1,i-1,j}^k}{2h} + \frac{g_{2,i,j+1}^k - g_{2,i,j-1}^k}{2h}.$$

The resulting linear system: $Au = r$

Finally, the resulting linear system $\boldsymbol{Au} = \boldsymbol{r}$ will be given by

$$
\begin{aligned}
&(\lambda + 4\frac{\gamma}{h^2})u_{i,j} - \frac{\gamma}{h^2}\Big(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}\Big) \\
&\quad = \lambda f_{i,j} - \frac{\gamma}{2h}\Big(g^k_{1,i+1,j} - g^k_{1,i-1,j} + g^k_{2,i,j+1} - g^k_{2,i,j-1}\Big), 1 \le i,j \le N.
\end{aligned}
$$

- Since $\lambda > 0$ and $\gamma > 0$, $\boldsymbol{Au} = \boldsymbol{r}$ will be symmetric and strictly diagonally dominant. It can be solved by many different methods such as the iterative techniques.

  Since $a_{ii} > 0$, we can prove that $\boldsymbol{A}$ is SPD by Gershgorin's Theorem!

- For example, the Gauss-Seidel iterative method gives

$$
\Big(\lambda + 4\frac{\gamma}{h^2}\Big)u^{k+1}_{i,j} = c^k_{i,j} + \frac{\gamma}{h^2}\Big(u^{k+1}_{i-1,j} + u^k_{i+1,j} + u^{k+1}_{i,j-1} + u^k_{i,j+1}\Big),\ k \ge 0,
$$

  where

$$
c^k_{i,j} := \lambda f_{i,j} - \frac{\gamma}{2h}\Big(g^k_{1,i+1,j} - g^k_{1,i-1,j} + g^k_{2,i,j+1} - g^k_{2,i,j-1}\Big).
$$

$\boldsymbol{d}$-subproblem

$\boldsymbol{d}$-subproblem: With u and $\boldsymbol{b}$ fixed, we solve

$$\boldsymbol{d}^{k+1} = \arg\min_{\boldsymbol{d}} \left(\sum_{i,j=1}^{N} |\boldsymbol{d}_{i,j}| + \frac{\gamma}{2} \sum_{i,j=1}^{N} |\boldsymbol{d}_{i,j} - \nabla_h u_{i,j}^{k+1} - \boldsymbol{b}_{i,j}^{k}|^2 \right),$$

which has a closed-form solution,

$$\boldsymbol{d}_{i,j}^{k+1} = \frac{\nabla_h u_{i,j}^{k+1} + \boldsymbol{b}_{i,j}^{k}}{|\nabla_h u_{i,j}^{k+1} + \boldsymbol{b}_{i,j}^{k}|} \max\left\{ |\nabla_h u_{i,j}^{k+1} + \boldsymbol{b}_{i,j}^{k}| - \frac{1}{\gamma},\ 0 \right\}, \quad 1 \le i,j \le N.$$

How to find the closed-form solution of $\boldsymbol{d}$-subproblem?

The solution of $\boldsymbol{d}$-subproblem can be found componentwisely. For each (i,j), we consider the following minimization problem:

$$\min_{\boldsymbol{x}=(x_1,x_2)^\top \in \mathbb{R}^2} \left\{ |\boldsymbol{x}| + \frac{\gamma}{2} |\boldsymbol{x} - \boldsymbol{c}|^2 \right\},$$

where $\gamma > 0$, $\boldsymbol{c} = (c_1, c_2)^\top \in \mathbb{R}^2$ are given, and $|\cdot| := \|\cdot\|_2$ in $\mathbb{R}^2$.

Solution of the d-subproblem

Suppose that $\boldsymbol{c} \neq \boldsymbol{0}$*, i.e.,* $|\boldsymbol{c}| > 0$*; otherwise the minimizer is* $\boldsymbol{x} = \boldsymbol{0}$*.* We consider the following minimization problem:

$$\min_{x_1,x_2}\left\{\sqrt{x_1^2+x_2^2}+\frac{\gamma}{2}\left((x_1-c_1)^2+(x_2-c_2)^2\right)\right\}.$$

With a careful inspection (by the triangle inequality), we can find that the minimizer will occur at $\boldsymbol{x} = t\boldsymbol{c}$ *for some* $t \in [0,1]$*, i.e.,* $x_1 = tc_1$ *and* $x_2 = tc_2$ *for some* $t \in [0,1]$*.* Therefore, the minimization problem can be rewritten as

$$\min_{t\in[0,1]}\underbrace{\left\{t\sqrt{c_1^2+c_2^2}+\frac{\gamma}{2}\left((t-1)^2c_1^2+(t-1)^2c_2^2\right)\right\}}_{:=g(t)}.$$

We can rewrite function g as

$$g(t) = t|\boldsymbol{c}|+\frac{\gamma}{2}(t-1)^2|\boldsymbol{c}|^2 = \frac{\gamma}{2}|\boldsymbol{c}|^2t^2+(|\boldsymbol{c}|-\gamma|\boldsymbol{c}|^2)t+\frac{\gamma}{2}|\boldsymbol{c}|^2, \quad t\in[0,1].$$

By direct computations, we have $g(0) = \frac{\gamma}{2}|\boldsymbol{c}|^2$, $g(1) = |\boldsymbol{c}|$, and

$$g'(t) = \gamma|\boldsymbol{c}|^2t+(|\boldsymbol{c}|-\gamma|\boldsymbol{c}|^2) \text{ for } t\in(0,1).$$

Solution of the d-subproblem (cont'd)

- If $|\boldsymbol{c}| > \frac{1}{\gamma} > 0$, we have a unique critical number of $g(t)$ in $(0,1)$,

$$t_0 = \frac{\gamma|\boldsymbol{c}|^2 - |\boldsymbol{c}|}{\gamma|\boldsymbol{c}|^2} = 1 - \frac{1}{\gamma|\boldsymbol{c}|} \in (0,1).$$

  Then $g(t_0) = |\boldsymbol{c}| - \frac{1}{2\gamma}$ is the minimum value and the minimizer is

$$\boldsymbol{x} = t_0\boldsymbol{c} = \left(1 - \frac{1}{\gamma|\boldsymbol{c}|}\right)\boldsymbol{c} = \left(|\boldsymbol{c}| - \frac{1}{\gamma}\right)\frac{\boldsymbol{c}}{|\boldsymbol{c}|}.$$

- If $0 < |\boldsymbol{c}| \le \frac{1}{\gamma}$, then $g(t)$ has no critical number in $(0,1)$ and $g(0) = \frac{\gamma}{2}|\boldsymbol{c}|^2 \le \frac{1}{2}|\boldsymbol{c}| \le |\boldsymbol{c}| = g(1)$. Therefore, $g(0)$ is the minimum of $g(t)$ on $[0,1]$ and $\boldsymbol{x} = \boldsymbol{0}$.

Combining these two cases, we have

$$\begin{aligned}\arg\min_{\boldsymbol{x}\in\mathbb{R}^2}\left\{|\boldsymbol{x}| + \frac{\gamma}{2}|\boldsymbol{x}-\boldsymbol{c}|^2\right\} &= \begin{cases} \dfrac{\boldsymbol{c}}{|\boldsymbol{c}|}\left(|\boldsymbol{c}| - \frac{1}{\gamma}\right), & \text{if } |\boldsymbol{c}| > \frac{1}{\gamma} \\ \boldsymbol{0}, & \text{if } |\boldsymbol{c}| \le \frac{1}{\gamma} \end{cases} \\ &= \frac{\boldsymbol{c}}{|\boldsymbol{c}|}\max\left\{|\boldsymbol{c}| - \frac{1}{\gamma},\ 0\right\}.\end{aligned}$$

Updating b and selecting γ

- **Updating $\boldsymbol{b}$:** The auxiliary variable $\boldsymbol{b}$ is initialized to zero and updated as

$$\boldsymbol{b}_{i,j}^{k+1} = \boldsymbol{b}_{i,j}^{k} + \nabla_h u_{i,j}^{k+1} - \boldsymbol{d}_{i,j}^{k+1}, \quad 1 \leq i,j \leq N.$$

- **Selecting γ:** A good choice of γ is one for which both u and $\boldsymbol{d}$ subproblems converge quickly and are numerically well-conditioned.
  - *In the u subproblem, the effect of $\nabla \cdot \nabla$ and $\nabla \cdot$ increase when γ gets larger. It is ill-conditioned in the limit $\gamma \to \infty$.*
  - In the $\boldsymbol{d}$ subproblem, the shrinking effect is more dramatic when γ is small.

Therefore, γ should be neither extremely large nor small for good convergence. In our simulations, we take $\gamma/h = 0.1$.

The split Bregman algorithm

The split Bregman algorithm:

initialize $u = f$, $\boldsymbol{d} = \boldsymbol{b} = \boldsymbol{0}$

while $\|u_{current} - u_{previous}\|_2 > tolerance$ *do*

solve the u*-subproblem*

solve the $\boldsymbol{d}$*-subproblem*

$\boldsymbol{b} = \boldsymbol{b} + \nabla u - \boldsymbol{d}$

Color images (RGB channels): The vectorial TV (VTV) is used in place of TV,

$$\|u\|_{VTV(\Omega)} := \int_\Omega \Big(\sum_{i \in \{R,G,B\}} |\nabla u_i(\boldsymbol{x})|^2 \Big)^{1/2} d\boldsymbol{x}.$$

The grayscale algorithm can be extended directly to VTV-regularized image denoising.

Implementation details of split Bregman iterations

u**-subproblem:** We multiply the following identity (see p. 43) with h,

$$\left(\lambda + 4\frac{\gamma}{h^2}\right)u_{i,j} - \frac{\gamma}{h^2}\Big(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}\Big)$$
$$= \lambda f_{i,j} - \frac{\gamma}{2h}\Big(g^k_{1,i+1,j} - g^k_{1,i-1,j} + g^k_{2,i,j+1} - g^k_{2,i,j-1}\Big), 1 \le i,j \le N.$$

Then we have

$$\left(\lambda h + 4\frac{\gamma}{h}\right)u_{i,j} - \frac{\gamma}{h}\Big(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}\Big)$$
$$= \lambda h f_{i,j} - \frac{\gamma}{2h}\Big(hg^k_{1,i+1,j} - hg^k_{1,i-1,j} + hg^k_{2,i,j+1} - hg^k_{2,i,j-1}\Big), 1 \le i,j \le N.$$

Notice that $g^k_{i,j} = (g^k_{1,i,j}, g^k_{2,i,j})^\top := \boldsymbol{d}^k_{i,j} - \boldsymbol{b}^k_{i,j}$. Define $\widetilde{\lambda} = \lambda h$, $\widetilde{\gamma} = \dfrac{\gamma}{h}$,

$\widetilde{g}^k_{i,j} = (\widetilde{g}^k_{1,i,j}, \widetilde{g}^k_{2,i,j})^\top := h\boldsymbol{d}^k_{i,j} - h\boldsymbol{b}^k_{i,j} := \widetilde{\boldsymbol{d}}^k_{i,j} - \widetilde{\boldsymbol{b}}^k_{i,j}$. Then we have

$$\left(\widetilde{\lambda} + 4\widetilde{\gamma}\right)u_{i,j} - \widetilde{\gamma}\Big(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}\Big)$$
$$= \widetilde{\lambda} f_{i,j} - \frac{\widetilde{\gamma}}{2}\Big(\widetilde{g}^k_{1,i+1,j} - \widetilde{g}^k_{1,i-1,j} + \widetilde{g}^k_{2,i,j+1} - \widetilde{g}^k_{2,i,j-1}\Big), 1 \le i,j \le N. \quad (\star_1)$$

Implementation details of split Bregman iterations (cont'd)

***d*-subproblem:** If we define

$$\widetilde{\nabla} u_{i,j} := [\delta_x u_{i,j}, \delta_y u_{i,j}]^\top := \left[\frac{u_{i+1,j} - u_{i-1,j}}{2}, \ \frac{u_{i,j+1} - u_{i,j-1}}{2}\right]^\top,$$

then since (see page 44)

$$\boldsymbol{d}_{i,j}^{k+1} = \frac{\nabla_h u_{i,j}^{k+1} + \boldsymbol{b}_{i,j}^k}{|\nabla_h u_{i,j}^{k+1} + \boldsymbol{b}_{i,j}^k|} \max\left\{|\nabla_h u_{i,j}^{k+1} + \boldsymbol{b}_{i,j}^k| - \frac{1}{\gamma}, \ 0\right\},$$

we have

$$\begin{aligned}
\widetilde{\boldsymbol{d}}_{i,j}^{k+1} &= h\boldsymbol{d}_{i,j}^{k+1} = \frac{h\nabla_h u_{i,j}^{k+1} + h\boldsymbol{b}_{i,j}^k}{|h\nabla_h u_{i,j}^{k+1} + h\boldsymbol{b}_{i,j}^k|} h \ \max\left\{|\nabla_h u_{i,j}^{k+1} + \boldsymbol{b}_{i,j}^k| - \frac{1}{\gamma}, \ 0\right\} \\
&= \frac{\widetilde{\nabla} u_{i,j}^{k+1} + \widetilde{\boldsymbol{b}}_{i,j}^k}{|\widetilde{\nabla} u_{i,j}^{k+1} + \widetilde{\boldsymbol{b}}_{i,j}^k|} \max\left\{|\widetilde{\nabla} u_{i,j}^{k+1} + \widetilde{\boldsymbol{b}}_{i,j}^k| - \frac{1}{\widetilde{\gamma}}, \ 0\right\}. \quad (\star_2)
\end{aligned}$$

Implementation details of split Bregman iterations (cont'd)

Updating b: First, we have (see page 47)

$$\boldsymbol{b}_{i,j}^{k+1} = \boldsymbol{b}_{i,j}^{k} + \nabla_h u_{i,j}^{k+1} - \boldsymbol{d}_{i,j}^{k+1}.$$

By multiplying the identity with h, we obtain

$$h\boldsymbol{b}_{i,j}^{k+1} = h\boldsymbol{b}_{i,j}^{k} + h\nabla_h u_{i,j}^{k+1} - h\boldsymbol{d}_{i,j}^{k+1}.$$

In other words,

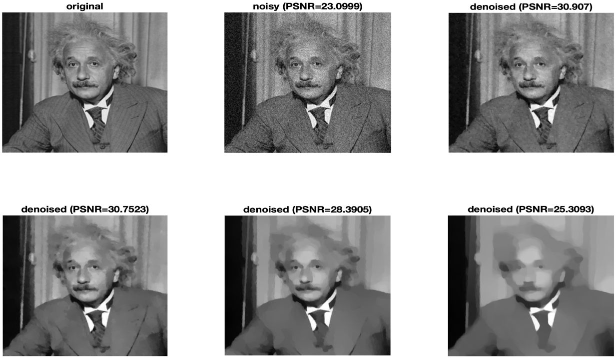
$$\widetilde{\boldsymbol{b}}_{i,j}^{k+1} = \widetilde{\boldsymbol{b}}_{i,j}^{k} + \widetilde{\nabla}_h u_{i,j}^{k+1} - \widetilde{\boldsymbol{d}}_{i,j}^{k+1}. \quad (\star_3)$$

Some remarks

To sum up, we have the following remarks:

- *By change of variables, the split Bregman iterations can be reformulated as* $(\star_1)$, $(\star_2)$, $(\star_3)$, *where the grid size h can be absorbed by other variables!*
- Most engineering-oriented papers usually take the spatial grid size $h = 1$ in the finite differences. It is irrational from the approximation viewpoint because the error terms in Taylor's theorem may not be small if we take $h = 1$.
- However, if the grid size h has been absorbed by other variables as discussed above, then it is reasonable for us to say that, in some sense, the grid size $h = 1$.

Numerical experiments (Einstein)



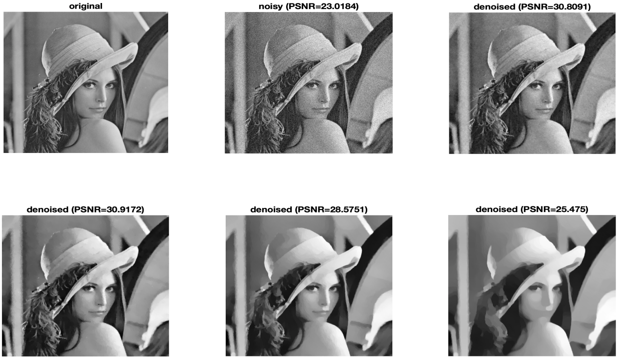
Gaussian noise $(0, 0.005)$, $h = 1/340$, $\tilde{\lambda} = h\lambda = 0.1, 0.05, 0.025, 0.01$, $\tilde{\gamma} = \gamma/h = 0.1$

A smaller value of λ implies stronger denoising. When λ is very small, the image becomes cartoon-like with sharp jumps between nearly flat regions.

Numerical experiments (Cameraman)

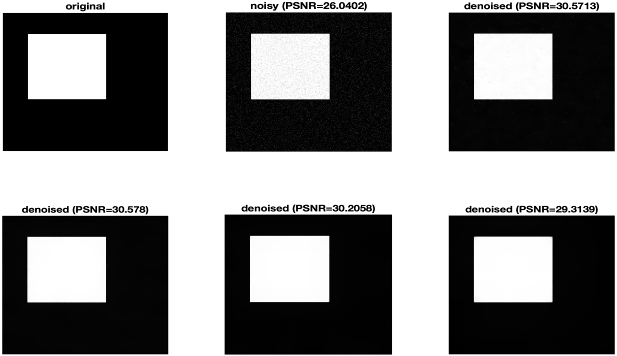
Gaussian noise $(0, 0.005)$, $h = 1/256$, $\widetilde{\lambda} = h\lambda = 0.1, 0.05, 0.025, 0.01$, $\widetilde{\gamma} = \gamma/h = 0.1$

Numerical experiments (Lena)



Gaussian noise $(0, 0.005)$, $h = 1/512$, $\tilde{\lambda} = h\lambda = 0.1, 0.05, 0.025, 0.01,$
$\tilde{\gamma} = \gamma/h = 0.1$

Numerical experiments (square)



Gaussian noise $(0, 0.005)$, $h = 1/256$, $\widetilde{\lambda} = h\lambda = 0.1, 0.05, 0.025, 0.01$,
$\widetilde{\gamma} = \gamma/h = 0.1$