Mathematics of Image and Data Analysis
Math 5467

Total Variation (TV) Denoising

Instructor: Jeff Calder
Email: jcalder@umn.edu

http://www-users.math.umn.edu/~jwcalder/5467

Last time

- Tikhonov regularized denoising

Today

- Total Variation (TV) regularized denoising

Tikhonov regularization

Let $f \in L^2(\mathbb{Z}_n)$ be the noisy signal. Tikhonov regularized denoising minimizes the energy $E : L^2(\mathbb{Z}_n) \to L^2(\mathbb{Z}_n)$ defined by

$$E(u) = \underbrace{\sum_{k=0}^{n-1} |u(k) - f(k)|^2}_{\text{Data Fidelity}} + \lambda \underbrace{\sum_{k=0}^{n-1} |u(k) - u(k-1)|^2}_{\text{Regularizer}}, \tag{1}$$

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

Main ideas:

- Data fidelity keeps the denoised signal close to the noisy signal f.
- Regularizer removes the noise.

Tikhonov regularization

We recall the backward difference $\nabla^- : L^2(\mathbb{Z}_n) \to L^2(\mathbb{Z}_n)$ is defined by

$$\nabla^- u(k) = u(k) - u(k-1),$$

while the forward difference is $\nabla^+ u(k) = u(k+1) - u(k)$. The discrete Laplacian is

$$\Delta u = \nabla^+ \nabla^- u = \nabla^- \nabla^+ u.$$

In terms of this notation, the Tikhonov regularized denoising problem is

$$\min_{u \in L^2(\mathbb{Z}_n)} E(u) = \|u - f\|^2 + \lambda \|\nabla^- u\|^2. \tag{2}$$

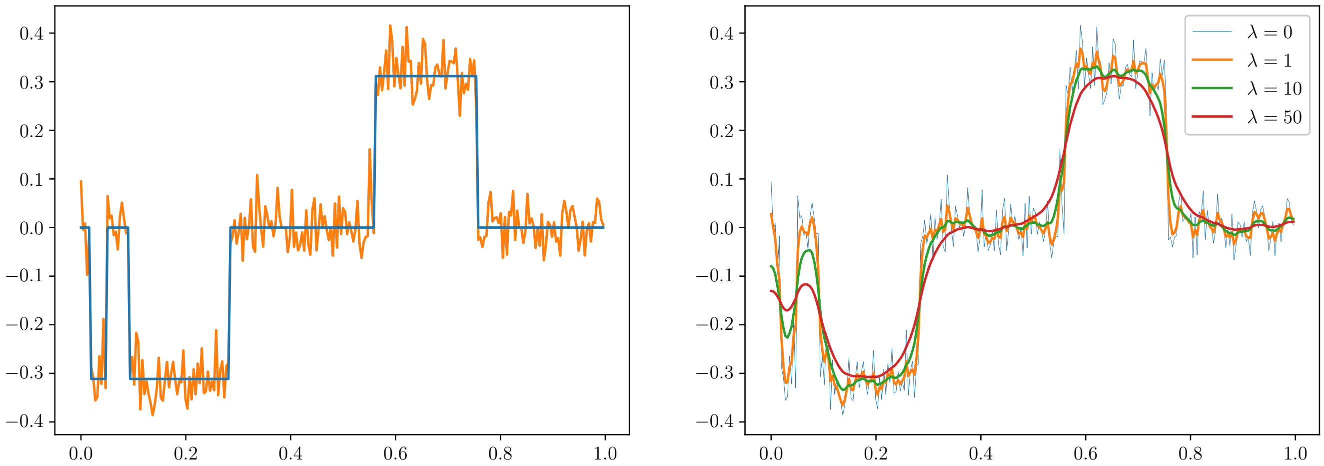
Theorem 1. *Let $\lambda \geq 0$ and $f \in L^2(\mathbb{Z}_n)$. Then there exists a unique solution $u \in L^2(\mathbb{Z}_n)$ of the optimization problem (2). Furthermore, the minimizer u is also characterized as the unique solution of the Euler-Lagrange equation*

$$u - \lambda \Delta u = f. \tag{3}$$

$\nabla E = 0$

Tikhonov regularization

Bar Code



(a) Noisy signal

(b) Tikhonov denoising

Total Variation Regularization

Total Variation (TV) regularization replaces the squared difference by the absolute differences in the regularizer.

$$E(u) = \frac{1}{2} \sum_{k=0}^{n-1} |u(k) - f(k)|^2 + \lambda \sum_{k=0}^{n-1} |u(k) - u(k-1)|. \tag{4}$$

Total Variation

- TV regularization is better at preserving edges (sharp changes) in the signal.
- The analysis is more involved, since the denosing equation is *nonlinear*.

Variational Regularized Denoising

$\|u\|_1 = \sum_{k=0}^{n-1} |u(k)|$

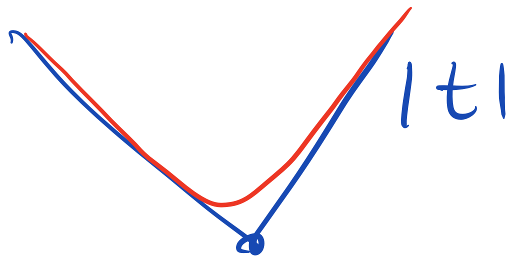
We will proceed in generality, studying regularizers of the form

$$\sum_{k=0}^{n-1} \Phi(u(k) - u(k-1)) = \sum_{k=0}^{n-1} \Phi(\nabla^- u(k)) = \|\Phi(\nabla^- u)\|_1, \tag{5}$$

where $\Phi : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable, convex, and even function satisfying $\Phi(0) = 0$.

not twice differentiable

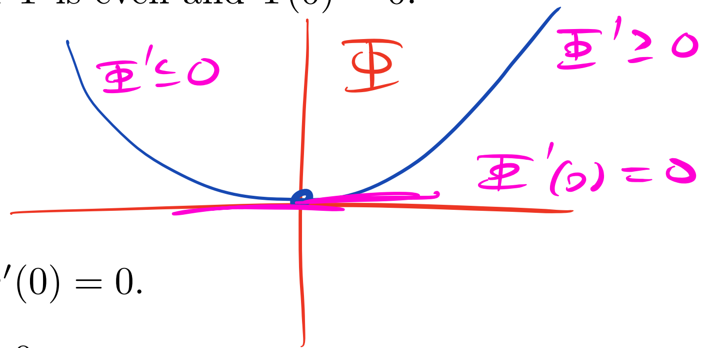
- Tikhonov is $\Phi(t) = t^2$
- Total Variation (TV) is $\Phi(t) = |t|$.
- We will approximate TV by $\Phi(t) = \sqrt{t^2 + \varepsilon^2}$.



Convexity

We say Φ convex if $\Phi'' \geq 0$. We also assumed Φ is even and $\Phi(0) = 0$.

The following properties hold:



(i) Φ' is increasing.

(ii) Since Φ is even and $\Phi(0) = 0$ we have $\Phi'(0) = 0$.

(iii) $\Phi'(t) \leq 0$ for $t < 0$ and $\Phi'(t) \geq 0$ for $t > 0$.

(iv) For any $t, s \in \mathbb{R}$ we have

$$(\Phi'(t) - \Phi'(s))(t - s) \geq 0.$$

$t \geq s$: ≥ 0 ≥ 0

$t \leq s$: ≤ 0 ≤ 0

Total Variation Denoising

The Total Variation (TV) regularized denoising function is

$$E_\Phi(u) = \frac{1}{2}\|u - f\|^2 + \lambda\|\Phi(\nabla^- u)\|_1. \tag{6}$$

The denoised signal u is found by minimizing E_Φ.

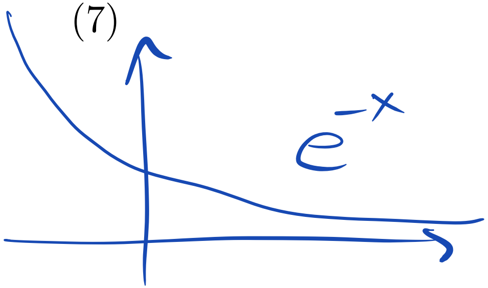
Note: We will work with real-value signals in this lecture, for simplicity. We denote by $L^2(\mathbb{Z}_n; \mathbb{R})$ the subspace of $L^2(\mathbb{Z}_n)$ consisting of $f : \mathbb{Z}_n \to \mathbb{R}$.

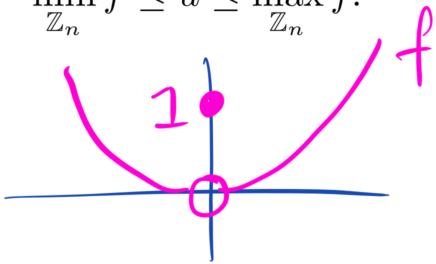
subspace of $L^2(\mathbb{Z}_n)$

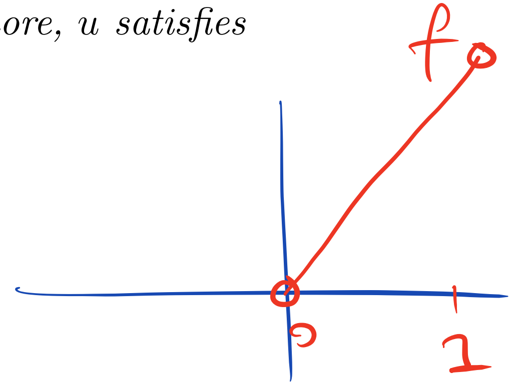
Existence of a minimizer

Lemma 2. *For any $f \in L^2(\mathbb{Z}_n;\mathbb{R})$ and $\lambda \geq 0$, there exists $u \in L^2(\mathbb{Z}_n;\mathbb{R})$ minimizing E_Φ, i.e., $E_\Phi(u) \leq E_\Phi(w)$ for all $w \in L^2(\mathbb{Z}_n;\mathbb{R})$. Furthermore, u satisfies*

$$\min_{\mathbb{Z}_n} f \leq u \leq \max_{\mathbb{Z}_n} f. \tag{7}$$





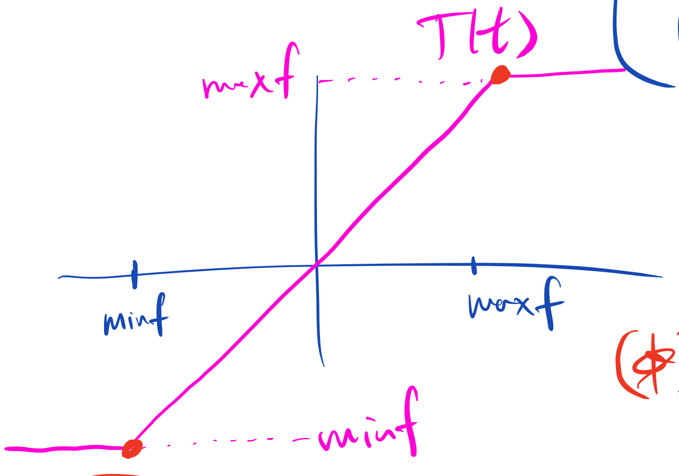


The proof is based on a simple fact: A continuous function on a closed and bounded subset of $\mathbb{R}^n$ attains its minimum value.

- $f(x) = e^{-x}$ does not have a minimum value on $\mathbb{R}$ (unbounded set).
- $f(x) = x^2$ for $x \neq 0$ and $f(0) = 1$ does not have a minimum value (discontinuous function).
- $f(x) = x$ does not have a minimum value on $(0, 1)$ (open set).

Proof of Lemma: E_Φ is continuous on $L^2(\mathbb{Z}_n,\mathbb{R})$ ($L^2(\mathbb{Z}_n,\mathbb{R}) \simeq \mathbb{R}^n$), but domain $\mathbb{R}^n$ is unbounded. Define truncation

$$T(t) = \begin{cases} \min f, & \text{if } t \le \min f \\ t, & \text{if } \min f \le t \le \max f \\ \max f, & \text{if } t \ge \max f. \end{cases}$$



Claim:

(*) $|T(t) - T(s)| \le |t - s|$.

$t \geq s$

$$|T(t) - T(s)| \leq \int_s^t |T'(\tau)| d\tau \leq t-s = |t-s|$$

($|T'(\tau)| \leq 1$)

Using claim, we'll show that

$$\boxed{E_\Phi(T(u)) \leq E_\Phi(u)}$$

To see this

$$\Phi(\nabla^- T(u))(k) = \Phi\big(T(u(k)) - T(u(k-1))\big)$$

Φ is even $\rightarrow$

$$= \Phi\big(|T(u(k)) - T(u(k-1))|\big)$$

... T is increasing on $[0,\infty)$

(*) + Φ is

$$\leq \Phi\left(|u(k) - u(k-1)| \right)$$

$$= \Phi(\nabla^{-} u)(k).$$

$$\|T(u) - f\|^2 = \sum_{k=0}^{n-1} |T(u(k)) - f(k)|^2$$

$$= \sum_{k=0}^{n-1} |T(u(k)) - T(f(k))|^2$$

(*)

$$\leq \sum_{k=0}^{n-1} |u(k) - f(k)|^2 = \|u - f\|^2$$

This proves the claim.

Thus, we can minimize E_{Φ} over the bounded set

$$A = \left\{ u \in L^2(\mathbb{Z}_n, \mathbb{R}) : \min f \le u \le \max f \right\}.$$

Thus a minimizer exists. ☒

Euler-Lagrange equation

Lemma 3. *Let $f \in L^2(\mathbb{Z}_n;\mathbb{R})$ and $\lambda \geq 0$. Then the minimizer $u \in L^2(\mathbb{Z}_n;\mathbb{R})$ of E_Φ is unique and is characterized as the unique solution of the Euler-Lagrange equation*

$$u - \lambda \nabla^+ \Phi'(\nabla^- u) = f. \tag{8}$$

$(\Phi(t) = \frac{1}{2}t^2)$

Δu for Tikhonov. $(\Phi'(t) = t)$

Recall

Proposition 4. *For all $u, v \in L^2(\mathbb{Z}_n)$ the following hold.*

(i) $\langle \nabla^- u, v \rangle = -\langle u, \nabla^+ v \rangle$

(ii) $\langle \nabla^+ u, v \rangle = -\langle u, \nabla^- v \rangle$

(iii) $\langle \Delta u, v \rangle = \langle u, \Delta v \rangle$

Proof: let u be a minimizer and take a variation of E_Φ in the direction $v \in L^2(\Omega, \mathbb{R})$:

$$e(t) = E_\Phi(u + tv)$$

Use that $e(t)$ has a minimum at $t = 0$, to get

$$e'(0) = 0$$

$$e'(t) = \frac{d}{dt} E_{\Phi}(u+tv)$$

$$= \frac{d}{dt}\left\{ \frac{\|u+tv-f\|^2}{2} + \lambda \sum_{k=0}^{n-1} \Phi\left(D^- u(k) + t D^- v(k)\right) \right\}$$

$$= \frac{d}{dt}\left\{ \frac{\|u-f\|^2}{2} + \frac{2t}{2}\langle u-f, v\rangle + \frac{t^2}{2}\|v\|^2 + \lambda \sum_{k=0}^{n-1} \Phi\left(D^- u(k) + t D^- v(k)\right) \right\}$$

$$= \cancel{2}\langle u-f, v\rangle + \cancel{2t\|v\|^2} + \lambda \sum_{k=0}^{n-1} \Phi'\left(D^- u(k) + \cancel{t D^- v(k)}\right) D^- v(k)$$

$$0 = e'(0) = \langle u - f, v \rangle + \lambda \langle \Phi'(\nabla^- u), \nabla^- v \rangle$$

Integrate by parts

$$= \langle u - f, v \rangle - \lambda \langle \nabla^+ \Phi'(\nabla^- u), v \rangle$$

$$e'(0) = \langle u - \lambda \nabla^+ \Phi'(\nabla^- u) - f, v \rangle = 0$$

$= 0$

for all $v \in L^2(\mathbb{Z}_n, \mathbb{R})$.

$$\Rightarrow \quad u - \lambda \nabla^+ \Phi'(\nabla^- u) = f \qquad (*)$$

This shows that every minimizer of E_Φ must satisfy (*).

Uniqueness: Let u, v solve (*).

$$u - \lambda \nabla^+ \Phi'(\nabla^- u) = f$$

$$-\left(v - \lambda \nabla^+ \Phi'(\nabla^- v) = f\right)$$

$$u - v - \lambda \nabla^+ \left(\Phi'(\nabla^- u) - \Phi'(\nabla^- v)\right) = 0$$

Take inner product with $u-v$ on both sides

$\langle u-v, u-v \rangle$

$$\|u-v\|^2 - \lambda \underbrace{\langle D^+(\Phi'(D^-u) - \Phi'(D^-v)), u-v \rangle}_{\text{integrate by parts}} = 0$$

$$\|u-v\|^2 + \lambda \underbrace{\langle \Phi'(D^-u) - \Phi'(D^-v), D^-u - D^-v \rangle}_{\geq 0} = 0$$

$$(\Phi'(t) - \Phi'(s))(t-s) \geq 0$$

by convexity of Φ

$$\Longrightarrow \|u-v\|^2 = 0 \Longrightarrow u = v$$

$\blacksquare$

The gradient of E_Φ

The gradient of E_Φ can be interpreted as

$$\nabla E_\Phi(u) = u - \lambda \nabla^+ \Phi'(\nabla^- u) - f. = 0$$

Recall $e(t) = E_\Phi(u+tv)$

$$e'(0) = \langle u - \lambda \nabla^+ \Phi'(\nabla^- u) - f, v \rangle$$

$$\left.\frac{d}{dt}\right|_{t=0} E_\Phi(u+tv) = \langle \underbrace{u - \lambda \nabla^+ \Phi'(\nabla^- u) - f}_{\nabla E_\Phi(u)}, v \rangle$$

Recall for $g: \mathbb{R}^n \to \mathbb{R}$,

$$\frac{d}{dt}\Big|_{t=0} g(x+ty) = Dg(x)\cdot y$$

↑ Chain rule

Definition of DE_Φ is

$$\frac{d}{dt}\Big|_{t=0} E_\Phi(u+tv) = \langle DE_\Phi(u), v\rangle$$

Gradient Descent

We can minimize E_Φ by gradient descent

$$u_{j+1} = u_j - dt \nabla E_\Phi(u) = u_j - dt \left(u_j - \lambda \nabla^+ \Phi'(\nabla^- u_j) - f \right)$$

Time step restriction: For stability and convergence of the gradient descent iteration, we have a time step restriction

$$dt \leq \frac{2}{1 + 4C_\Phi \lambda},$$

where $C_\Phi = \max_{t \in \mathbb{R}} \Phi''(t)$. This follows from a Von Nuemann analysis using the DFT.

First study $u_{j+1} = u_j - dt(u_j - c\lambda \Delta u)$

$$u_{j+1} = (1-dt)u_j + C\,dt\,\lambda\,\Delta u$$

Take DFT on both sides

$$D(u_{j+1}) = (1-dt)D(u_j) + C\,dt\,\lambda\,D(\Delta u_j)$$

$$\mathcal{D}(\Delta f)(k) = (\omega^k + \omega^{-k} - 2)\mathcal{D}f(k) = 2(\cos(2\pi k/n) - 1)\mathcal{D}f(k).$$

$$D(u_{j+1}) = \left[1 - dt + 2C\,dt\,\lambda\left(\cos\left(\tfrac{2\pi k}{n}\right) - 1\right)\right]D(u_j)$$

For stability we need

$$-1 \leq \underbrace{1 - dt + 2C\,dt\,\lambda\left(\cos\left(\tfrac{2\pi k}{n}\right) - 1\right)}_{A} \leq 1$$

$\cos = 1$ $\quad A \le 1 - dt$ need ≤ 1 ✓

Need

$\cos = -1$ $\quad A \ge 1 - dt - 4C\,dt\,\lambda \ge -1$

$$1 - (4C\lambda + 1)dt \ge -1$$

$$(4C\lambda + 1)dt \le 2$$

$$dt \le \frac{2}{4C\lambda + 1}$$

CFL condition

Von Neumann Analysis

For the real equation, we approximate

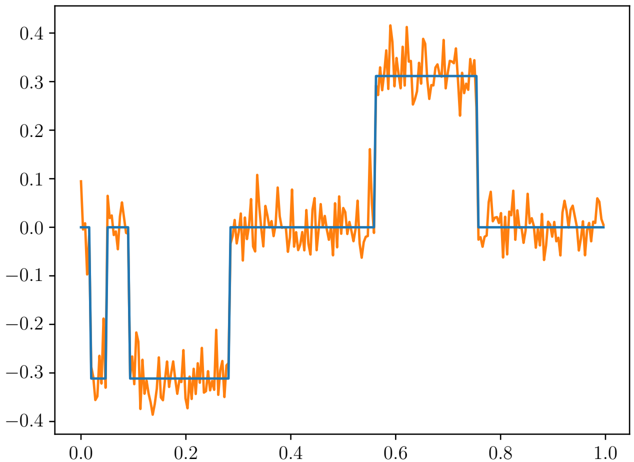
$$\nabla^+ \Phi'(\nabla^- u(k)) = \Phi'(\nabla^- u(k)) - \Phi'(\nabla^- u(k-1))$$

$$\simeq \Phi''(\nabla^- u(k))\,(\nabla^- u(k) - \nabla^- u(k-1))$$

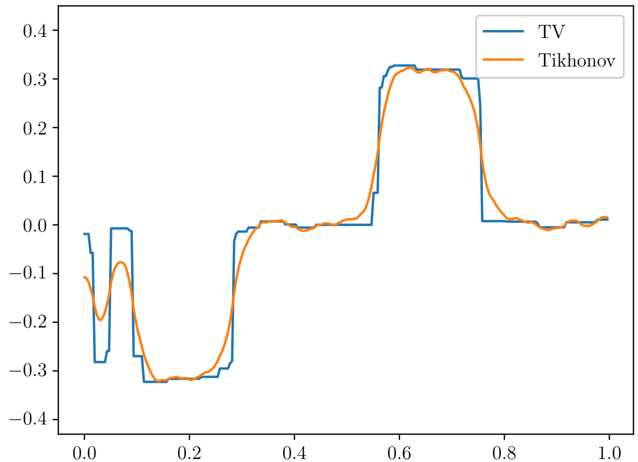
$$= \Phi''(\nabla^- u)\,\Delta u.$$

$$C_\Phi = \max |\Phi''|$$

Total Variation Denoising



(c) Noisy signal



(d) Denoised

Convergence of Gradient Descent

Theorem 5. *Let* $f \in L^2(\mathbb{Z}_n; \mathbb{R})$ *and* $\lambda \geq 0$. *Let* u_j *be the iterations of the gradient descent scheme for minimizing* E_Φ *and let* u *be the solution of* (8) *(the minimizer of* E_Φ*). Assume that the time step* dt *satisfies*

$$dt < \frac{2}{1 + 16 C_\Phi^2 \lambda^2}. \tag{9}$$

Then u_j *converges to* u *as* $j \to \infty$, *and the difference* $u_j - u$ *satisifes*

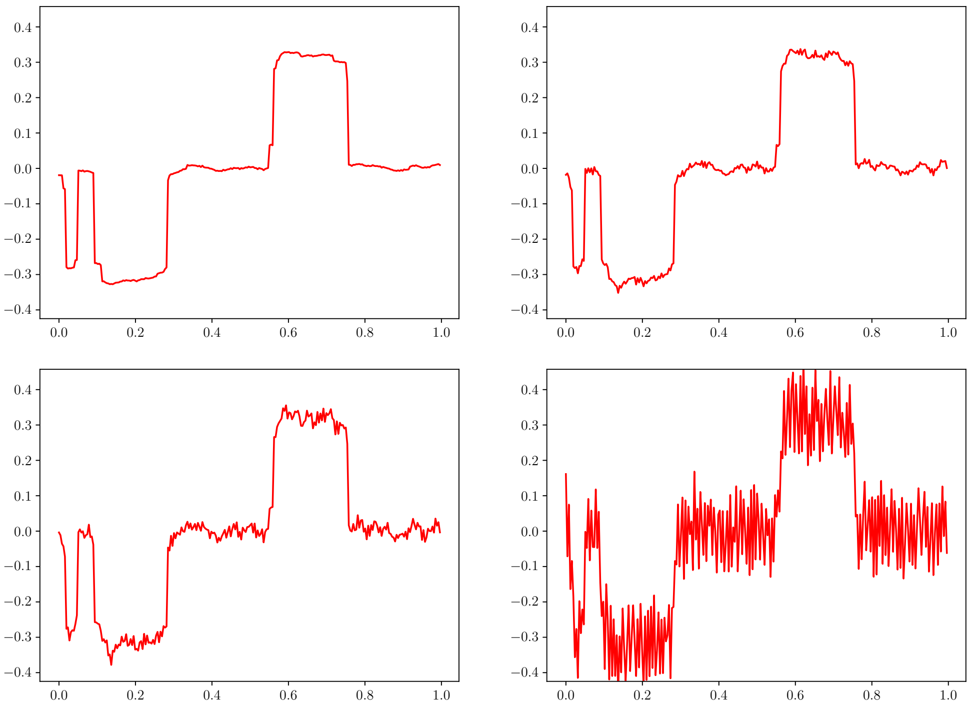
$$\|u_{j+1} - u\|^2 \leq \mu \|u_j - u\|^2 \tag{10}$$

where

$$\mu := (1 - dt)^2 + 16 C_\Phi^2 dt^2 \lambda^2 < 1. \tag{11}$$

Nonlinear stability at larger time steps

We set $\varepsilon = 10^{-10}$ and the CFL condition is $dt \sim 5 \times 10^{-10}$.
Figures are $dt = 0.01, 0.05, 0.1, 0.5$.



Local nonlinear stability

A heuristic local version of the Von Neumann analysis for ε-regularzied TV shows that the scheme is stable wherever the gradient of u satisfies

$$|\nabla^- u|^3 \geq \frac{4\lambda\varepsilon^2 dt}{2 - dt}.$$

Thus, oscillations cannot grow infinitely large, since the scheme is stable for larger gradients.

Total Variation denoising (.ipynb)