Variational Methods in CT Reconstruction

Chapter 12.4 and 12.5

Yiqiu Dong

DTU Compute
Technical University of Denmark

18 Jan. 2023

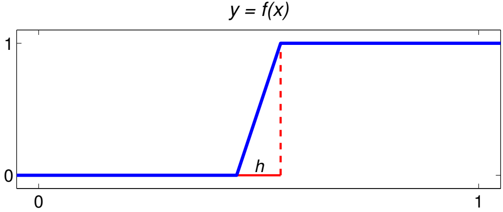
MAP estimate

The **MAP estimation** problem is

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2 + \alpha\, J(\boldsymbol{x}).$$

- The term $\frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2$ is called the *data-fidelity* term.
- The term $J(\boldsymbol{x})$ is called the *regularization* term.
- $\alpha > 0$ is the regularization parameter.
- *Tikhonov regularization*: $J(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|_2^2$.
- *Tikhonov regularization in general form*: $J(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{D}\boldsymbol{x}\|_2^2$.

An example using a continuous function



Consider the piecewise linear function

$$f(t) = \begin{cases} 0, & 0 \le t < \frac{1}{2}(1-h) \\ \dfrac{t}{h} - \dfrac{1-h}{2h}, & \frac{1}{2}(1-h) \le t \le \frac{1}{2}(1+h) \\ 1, & \frac{1}{2}(1+h) < t \le 1 \end{cases}$$

which increases linearly from 0 to 1 in $[\frac{1}{2}(1-h), \frac{1}{2}(1+h)]$.

Norms of the First Derivative

It is easy to show that the 1- and 2-norms of $f'(t)$ satisfy

$$\begin{aligned}\|f'\|_1 &= \int_0^1 |f'(t)|\,dt = \int_0^h \frac{1}{h}\,dt = 1,\\ \|f'\|_2^2 &= \int_0^1 f'(t)^2\,dt = \int_0^h \frac{1}{h^2}\,dt = \frac{1}{h}.\end{aligned}$$

Note that $\|f'\|_1$ is independent of the slope of the middle part of $f(t)$, while $\|f'\|_2$ penalizes steep gradients (when h is small).

- The 2-norm of $f'(t)$ will not allow any steep gradients and therefore it produces a smooth solution .
- The 1-norm, on the other hand, allows some steep gradients – but not too many – and it is therefore able to produce a less smooth solution, and even a discontinuous solution.

Total Variation (TV) Regularization

- **In 1D:**

$$\mathrm{TV}(\boldsymbol{x}) = \|\boldsymbol{D}_n \boldsymbol{x}\|_1 \ ,$$

where $\|\boldsymbol{x}\|_1$ denotes 1-norm of vector $\boldsymbol{x}$ defined as $\|\boldsymbol{x}\|_1 = |x_1| + \cdots + |x_n|$.

- **In 2D:**

$$\begin{aligned} \mathrm{TV_a}(\boldsymbol{x}) &= \|(\boldsymbol{I}_N \otimes \boldsymbol{D}_M)\boldsymbol{x}\|_1 + \|(\boldsymbol{D}_N \otimes \boldsymbol{I}_M)\boldsymbol{x}\|_1 \\ &= \sum_{i=1}^{n} \left(\left|[(\boldsymbol{I}_N \otimes \boldsymbol{D}_M)\boldsymbol{x}]_i\right| + \left|[(\boldsymbol{D}_N \otimes \boldsymbol{I}_M)\boldsymbol{x}]_i\right| \right) \ , \end{aligned}$$

$$\mathrm{TV_i}(\boldsymbol{x}) = \sum_{i=1}^{n} \sqrt{[(\boldsymbol{I}_N \otimes \boldsymbol{D}_M)\boldsymbol{x}]_i^2 + [(\boldsymbol{D}_N \otimes \boldsymbol{I}_M)\boldsymbol{x}]_i^2} \ ,$$

where $[\boldsymbol{z}]_i$ denotes the ith element of the vector $\boldsymbol{z}$.

Total Variation (TV) Regularization (2D)

$$\begin{aligned}\mathrm{TV_a}(\boldsymbol{x}) &= \|(\boldsymbol{I}_N \otimes \boldsymbol{D}_M)\boldsymbol{x}\|_1 + \|(\boldsymbol{D}_N \otimes \boldsymbol{I}_M)\boldsymbol{x}\|_1 \\ &= \sum_{i=1}^{n} \left(\left|[(\boldsymbol{I}_N \otimes \boldsymbol{D}_M)\boldsymbol{x}]_i\right| + \left|[(\boldsymbol{D}_N \otimes \boldsymbol{I}_M)\boldsymbol{x}]_i\right| \right) \ , \\ \mathrm{TV_i}(\boldsymbol{x}) &= \sum_{i=1}^{n} \sqrt{[(\boldsymbol{I}_N \otimes \boldsymbol{D}_M)\boldsymbol{x}]_i^2 + [(\boldsymbol{D}_N \otimes \boldsymbol{I}_M)\boldsymbol{x}]_i^2} \ .\end{aligned}$$

- The vector

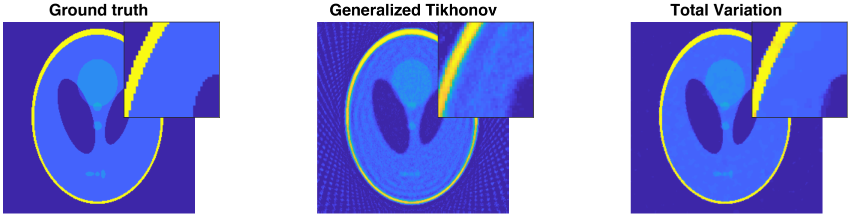
$$\begin{pmatrix} [(\boldsymbol{I}_N \otimes \boldsymbol{D}_M)\boldsymbol{x}]_i \\ [(\boldsymbol{D}_N \otimes \boldsymbol{I}_M)\boldsymbol{x}]_i \end{pmatrix}$$

  represents the gradient at the ith pixel in $\boldsymbol{x}$.
- In $\mathrm{TV_a}$, we use the 1-norm of the gradient for each pixel.
- In $\mathrm{TV_i}$, we use the 2-norm of the gradient for each pixel.

TV regularized CT reconstruction problem

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2 + \alpha\,\mathrm{TV}_{\square}(\boldsymbol{x}) ,$$

where $\square$ stands for either "a" or "i".



- By allowing occasional larger jumps in the reconstruction, it leads to piecewise constant results with sharp edges.
- The price is that the problem is non-differentiable.

Total Variation in Continuous Setting (1D)

The total variation of $f : [a, b] \to \mathbb{R}$ is defined as

$$TV(f) = \sup_{p \in \mathcal{P}} \sum_{i=1}^{n_p} |f(t_i) - f(t_{i-1})|,$$

where $\mathcal{P} = \{p = \{t_0, \cdots, t_{n_p}\} | p \text{ is a partition of } [a, b]\}$. The supremum is taken over all partitions of the interval $[a, b]$.

- If f is differentiable, then $TV(f) = \int_a^b |f'(t)|\, dt$
- Example: If $f(t) = \begin{cases} -1, & \text{if } -1 \leq t \leq 0, \\ 1, & \text{if } 0 < t \leq 1 \end{cases}$, then $TV(f) = 2$.
- Bounded variation (BV) space is defined as

$$f \in BV([a, b]) \iff TV(f) < +\infty.$$

- $W^{1,1}([a, b]) \subset BV([a, b]) \subset L^1([a, b])$.
- In discrete case, if f is differentiable, then $TV(f) = \sum_{i=1}^{n} |f'(t_i)|$.

Sparse regularization

Idea: Look for a "sparse" solution with many zero elements.

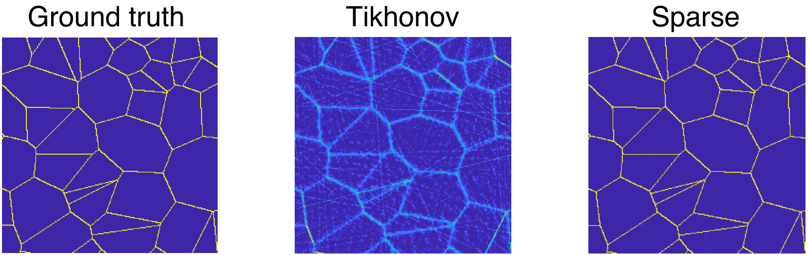
$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2 + \alpha\,\|\boldsymbol{x}\|_0$$

- $\|\boldsymbol{x}\|_0 = \sharp\{i = 1, \cdots, n : x_i \neq 0\}$.
- It is computational very challenging.
- The 0-norm is often approximated by
  - the q-norm, $\|\boldsymbol{x}\|_q = (|x_1|^q + \cdots + |x_n|^q)^{1/q}$, with $0 < q < 1$ (from bridge regression);
  - the 1-norm (most commonly used).

Sparse regularization

Idea: Look for a “sparse” solution with many zero elements.

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2 + \alpha\,\|\boldsymbol{x}\|_1 \qquad \text{(Lasso)}$$



- The minimization problem is non-differentiable.
- The corresponding prior probability density is the Laplace distribution.

Sparse Regularization

In practice, sparsity regularization is often applied with respect to a certain orthonormal basis.

We consider an orthonormal basis $\{\psi_s\}$, e.g. wavelet basis, then formally we have

$$\langle \boldsymbol{x}, \psi_s \rangle = \langle \boldsymbol{A}^\dagger \boldsymbol{b}, \psi_s \rangle = \langle \bar{\boldsymbol{x}}, \psi_s \rangle + \langle \boldsymbol{A}^\dagger \boldsymbol{e}, \psi_s \rangle.$$

If the basis $\{\psi_s\}$ is efficient to represent $\bar{\boldsymbol{x}}$, i.e., the most of the coefficients $\langle \bar{\boldsymbol{x}}, \psi_s \rangle$ are close to zero. But noise leads to a large set of non-zero coefficients, so we can use the sparse regularization in order to obtain sparse coefficients.

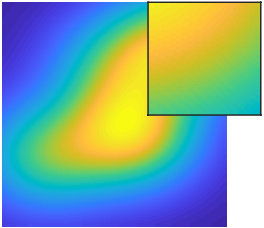
$$\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \alpha \sum_s \|\langle \boldsymbol{x}, \psi_s \rangle\|_1.$$

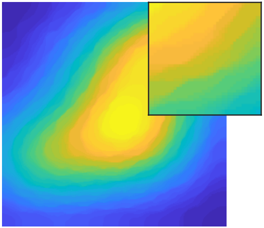
Stair-casing artifacts from TV

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_2^2 + \alpha\,\mathrm{TV}_{\square}(\boldsymbol{x})$$

Due to the relation between the sparsity and the 1-norm, TV regularization can be interpreted as sparsity regularization on the gradient, which leads to a piecewise constant solution.

- **Pro:** It can reconstruct sharp edges.
- **Con:** Smoothly varying parts in $\bar{\boldsymbol{x}}$ are often approximated by piecewise constant structures. This phenomena is called stair-casing artifacts.





Exercise

12.7 **Total Variation for a 2D Function**

12.8 **Numerical Computation of TV**