Characterizing the maximum parameter of the total-variation denoising through the pseudo-inverse of the divergence

Charles Deledalle[1], Nicolas Papadakis[1], Joseph Salmon[2] and Samuel Vaiter[3]

[1]IMB, CNRS Université de Bordeaux — [2]LTCI, Télécom ParisTech, Université Paris-Saclay — [3]IMB, CNRS Université de Bourgogne

Context

- Find the maximum regularization parameter for anisotropic total-variation denoising
- ≡ minimum value above which the solution remains constant
- well known for the Lasso,
  but, not yet investigated in details for the total-variation
- important when tuning the regularization parameter,
  provides an upper-bound on the grid for which the optimal parameter is sought

Contributions

- Closed form expression for the one-dimensional case
- upper-bound for the two-dimensional case,
  appears reasonably tight in practice
- computation of the pseudo-inverse of the divergence,
  quickly obtained by performing convolutions in the Fourier domain

Problem statement

- Anisotropic TV regularization writes, for $\lambda > 0$, as [Rudin et al., 1992]

$$x^\star = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\|y - x\|_2^2 + \lambda \|\nabla x\|_1 \quad (1)$$

- $y = x + w \in \mathbb{R}^n$: a noisy observation with $w \in \mathbb{R}^n$
- $x \in \mathbb{R}^n$: a d-dimensional signal (in this study $d = 1$ or 2)
- $\nabla x \in \mathbb{R}^{dn}$: discrete periodical gradient vector field of x
- $\|\nabla x\|_1 = \sum_i |(\nabla x)_i|$: a gradient-sparsity promoting term

Gradient / divergence / Fourier domain

∇ acts as a convolution which writes in the one dimensional case ($d = 1$)

$$\nabla = F^+ \operatorname{diag}(K_\downarrow) F \quad \text{and} \quad \operatorname{div} = -\nabla^\top = F^+ \operatorname{diag}(K_\uparrow) F \quad (2)$$

- $^\top$: denotes the adjoint
- $F : \mathbb{R}^n \mapsto \mathbb{C}^n$: the discrete Fourier transform
- $F^+ = \text{Re}[F^{-1}]$: its pseudo-inverse
- $K_\downarrow \in \mathbb{C}^n$ & $K_\uparrow \in \mathbb{C}^n$: Fourier transforms of the kernel functions performing forward and backward finite differences respectively

Similarly, we define in the two dimensional case ($d = 2$)

$$\nabla = \begin{pmatrix} F^+ & 0 \\ 0 & F^+ \end{pmatrix} \begin{pmatrix} \operatorname{diag}(K_\downarrow) \\ \operatorname{diag}(K_\rightarrow) \end{pmatrix} F \quad (3)$$

$$\text{and} \quad \operatorname{div} = F^+ \left(\operatorname{diag}(K_\uparrow) \ \operatorname{diag}(K_\leftarrow) \right) \begin{pmatrix} F & 0 \\ 0 & F \end{pmatrix} \quad (4)$$

- $K_\rightarrow \in \mathbb{C}^n$ & $K_\leftarrow \in \mathbb{C}^n$: forward and backward finite difference in horizontal direction
- $K_\downarrow \in \mathbb{C}^n$ & $K_\uparrow \in \mathbb{C}^n$: forward and backward finite difference in vertical direction

Proposition (General case)

- *Define for $y \in \mathbb{R}^n$,*

$$\lambda_{\max} = \min_{\zeta \in \operatorname{Ker}[\operatorname{div}]} \|\operatorname{div}^+ y + \zeta\|_\infty \quad (5)$$

- *Then,* $$x^\star = \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\top y \quad \textit{if and only if} \quad \lambda \geqslant \lambda_{\max} \quad (6)$$

- div^+: *Moore-Penrose pseudo-inverse of* div
- $\operatorname{Ker}[\operatorname{div}]$: *null space of* div
- *Proof:* *direct consequence of the Karush-Khun-Tucker condition*
- *Remark:* *non-smooth convex optimization problem*

Corollary (Mono dimensional case)

- *For* $d = 1$, $\lambda_{\max} = \frac{1}{2}[\max(\operatorname{div}^+ y) - \min(\operatorname{div}^+ y)]$,

$$\textit{where} \quad \operatorname{div}^+ = F^+ \operatorname{diag}(K_\uparrow^+) F \quad \textit{and} \quad (K_\uparrow^+)_i = \begin{cases} \frac{(K_\uparrow)_i^*}{|(K_\uparrow)_i|^2} & \textit{if} \quad |(K_\uparrow)_i|^2 > 0 \\ 0 & \textit{otherwise} \end{cases},$$

- *: *complex conjugatex*
- *Proof:* *since in the 1d case,* $\operatorname{Ker}[\operatorname{div}] = \operatorname{Span}(\mathbb{1}_n)$
- *Consequence:* $O(n \log n)$ *using the Fast Fourier Transform (FFT)*
- *Remark 1:* $|(K_\uparrow)_i|^2 > 0$ *everywhere except for the zero frequency*
- *Remark 2:* *non-periodical case,* div *is the incidence matrix of a tree whose pseudo-inverse is obtained following [Bapat, 1997]*

Difficulty in the two dimensional case

- $\operatorname{Ker}[\operatorname{div}]$: orthogonal of the vector space of vector fields satisfying Kirchhoff's voltage law on all cycles of the periodical grid
- $\dim \operatorname{Ker}[\operatorname{div}] = n + 1$ ⇒ optimization problem becomes much harder
  ⇒ resort to a fast approximation

Corollary

- *For* $d = 2$, $\lambda_{\max} \leqslant \underbrace{\frac{1}{2}[\max(\operatorname{div}^+ y) - \min(\operatorname{div}^+ y)]}_{\lambda_{\text{bnd}}}$,

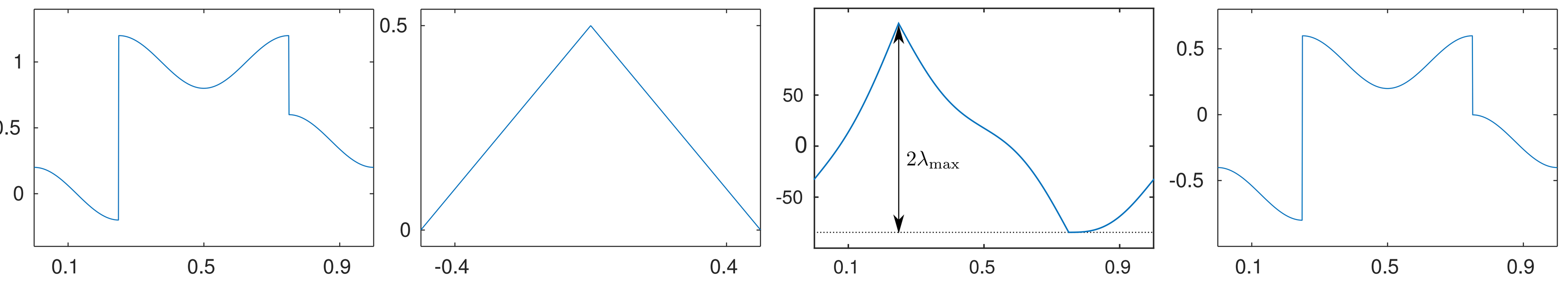
$$\textit{where} \quad \operatorname{div}^+ = \begin{pmatrix} F^+ & 0 \\ 0 & F^+ \end{pmatrix} \begin{pmatrix} \operatorname{diag}(\tilde{K}_\uparrow^+) \\ \operatorname{diag}(\tilde{K}_\leftarrow^+) \end{pmatrix} F, \quad \textit{and} \quad (7)$$

$$(\tilde{K}_\uparrow^+)_i = \begin{cases} \frac{(K_\uparrow)_i^*}{|(K_\uparrow)_i|^2 + |(K_\leftarrow)_i|^2} & \textit{if} \quad |(K_\uparrow)_i|^2 + |(K_\leftarrow)_i|^2 > 0 \\ 0 & \textit{otherwise} \end{cases},$$

$$(\tilde{K}_\leftarrow^+)_i = \begin{cases} \frac{(K_\leftarrow)_i^*}{|(K_\uparrow)_i|^2 + |(K_\leftarrow)_i|^2} & \textit{if} \quad |(K_\uparrow)_i|^2 + |(K_\leftarrow)_i|^2 > 0 \\ 0 & \textit{otherwise} \end{cases}.$$

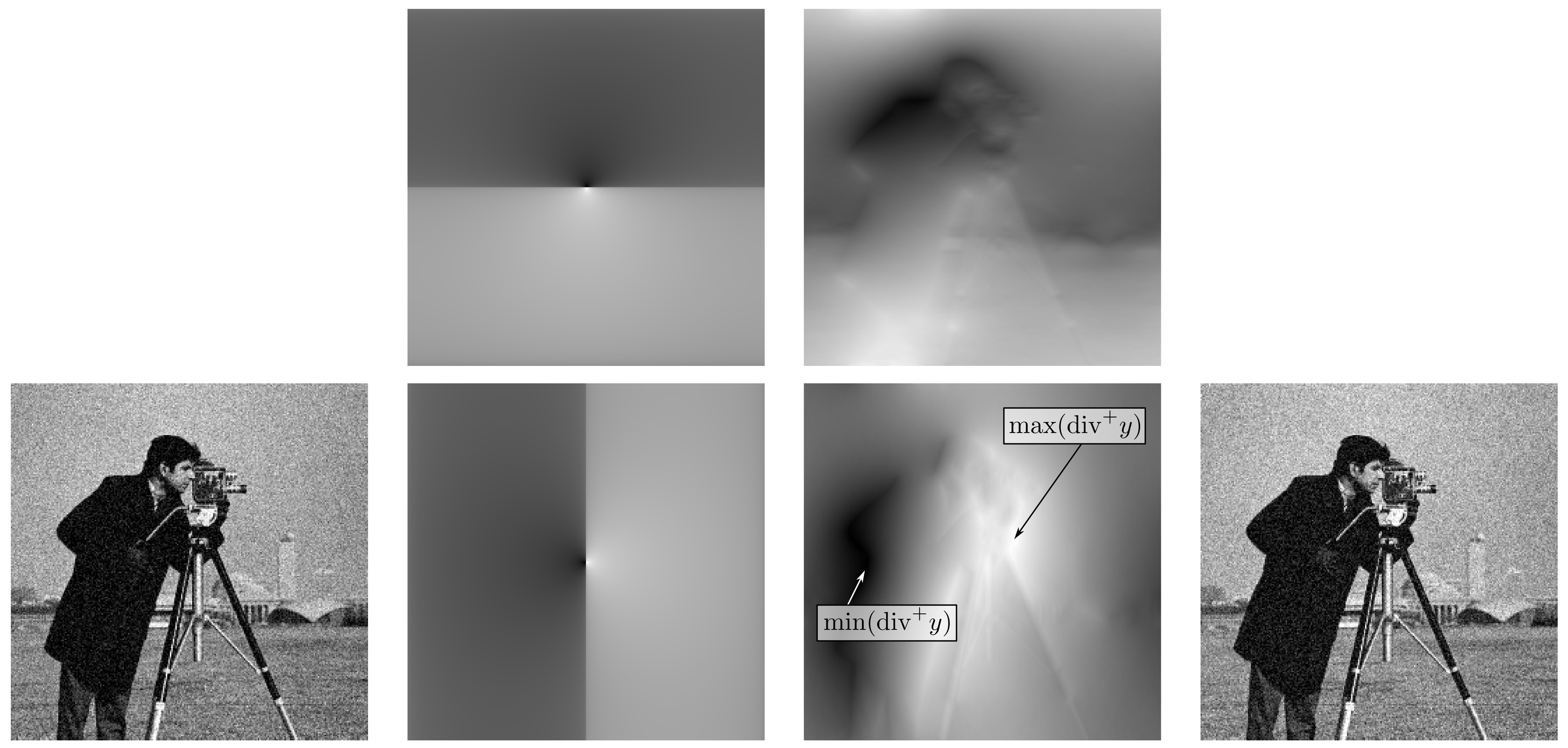
- *Proof:* *by direct calculus*
- *Consequence:* $O(n \log n)$ *operations using the 2D FFT*
- *Remark 1:* $|(K_\uparrow)_i|^2 + |(K_\leftarrow)_i|^2 > 0$ *except for the zero frequency*
- *Remark 2:* *can be straightforwardly extended to the case where* $d > 2$

Results and discussion



(a) y (b) $F^+ K_\uparrow^+$ (c) $\operatorname{div}^+ y$ (d) $\operatorname{div} \operatorname{div}^+ y$

Figure: (a) A 1d signal y. (b) The convolution kernel $F^+ K_\uparrow^+$ that realizes the pseudo inversion of the divergence. (c) The signal $\operatorname{div}^+ y$ on which we can read the value of $\lambda_{\max}$. (d) The signal $\operatorname{div} \operatorname{div}^+ y$ showing that one can reconstruct y from $\operatorname{div}^+ y$ up to its mean component.



(a) y (b) $F^+ \tilde{K}_\uparrow^+$, $F^+ \tilde{K}_\leftarrow^+$ (c) $\operatorname{div}^+ y$ (d) $\operatorname{div} \operatorname{div}^+ y$

Figure: (a) A 2d signal y. (b) The convolution kernels $F^+ K_\uparrow^+$ and $F^+ \tilde{K}_\leftarrow^+$ that realizes the pseudo inversion of the divergence. (c) The vector field $\operatorname{div}^+ y$ on which we can read the upper-bound λ_{bnd} of $\lambda_{\max}$. (d) The image $\operatorname{div} \operatorname{div}^+ y$ showing that one can reconstruct y from $\operatorname{div}^+ y$ up to its mean component.

(a) (b) $\lambda = 10^{-3} \lambda_{\text{bnd}}$ (c) $\lambda = 10^{-2} \lambda_{\text{bnd}}$ (d) $\lambda = \lambda_{\text{bnd}}$

Figure: (a) Evolution of $\|\nabla x^\star\|_\infty$ as a function of λ. (b), (c), (d) Results $x^\star$ of the periodical anisotropic total-variation for three different values of λ.

- Convolution kernel: simple triangle wave in 1d, but more complex in 2d
- $\operatorname{div} \operatorname{div}^+$ is the projector onto the space of zero-mean signals, i.e., $\operatorname{Im}[\operatorname{div}]$
- λ_{bnd}: computed in $\sim$5ms
- $\lambda_{\max}$: computed in $\sim$25s with [Chambolle and Pock, 2011] on Problem (5)
- λ_{bnd} appears to be reasonably tight upper bound of $\lambda_{\max}$
- Future work: other ℓ_1 sparse analysis regularization and ill-posed inverse problems

References

Bapat, R. (1997).
Moore-penrose inverse of the incidence matrix of a tree.
Linear and Multilinear Algebra, 42(2):159–167.

Chambolle, A. and Pock, T. (2011).
A first-order primal-dual algorithm for convex problems with applications to imaging.
J. Math. Imaging Vis., 40:120–145.

Rudin, L. I., Osher, S., and Fatemi, E. (1992).
Nonlinear total variation based noise removal algorithms.
Physica D: Nonlinear Phenomena, 60(1):259–268.