Higher-Order Total Variation Classes on Grids: Minimax Theory and Trend Filtering Methods

Veeranjaneyulu Sadhanala, Yu-Xiang Wang, James Sharpanack, Ryan J. Tibshirani

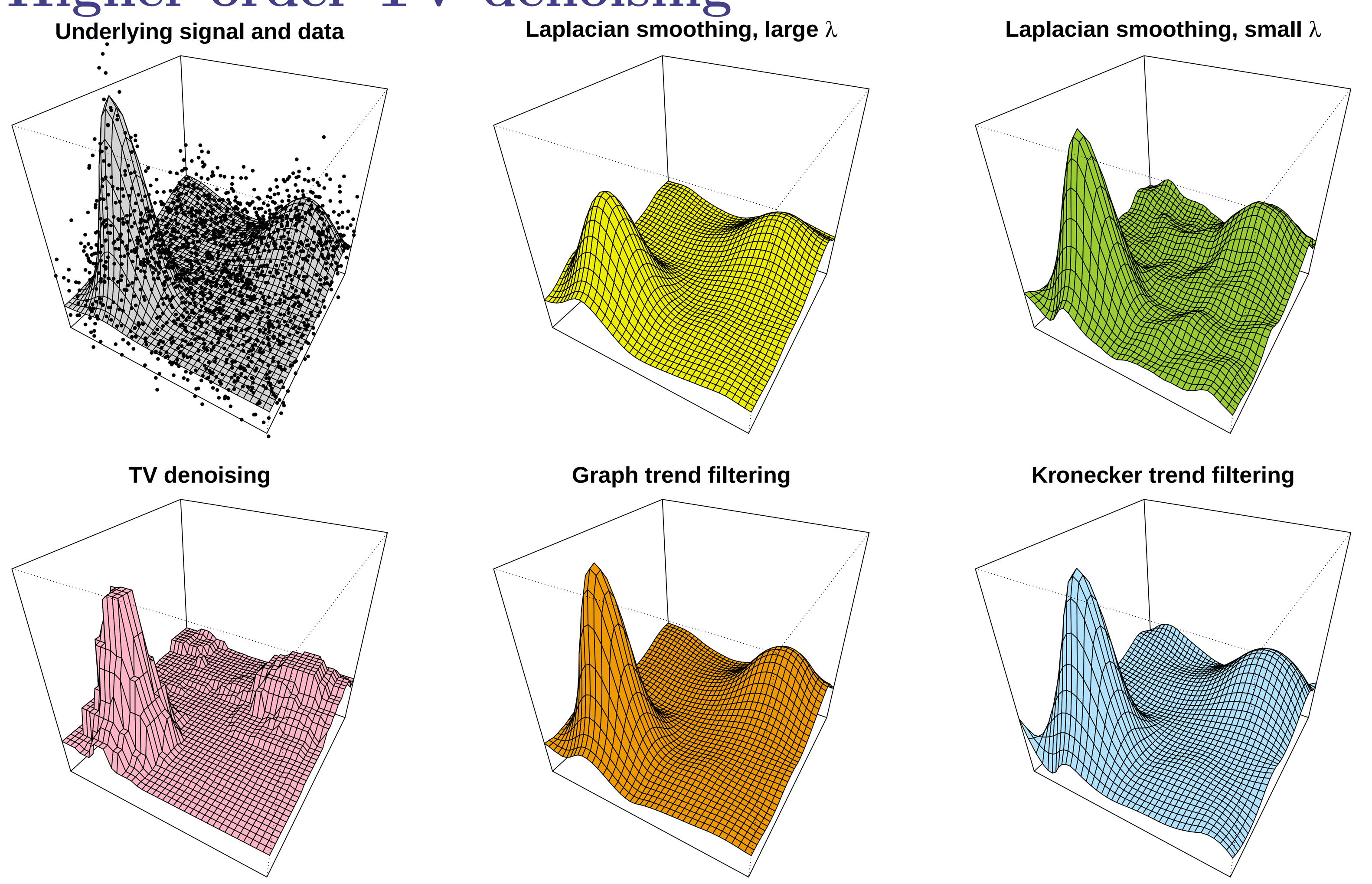
Carnegie Mellon University — Machine Learning Department

amazon web services

UC Davis Department of Statistics

Introduction and Objective

◇ Higher-order TV denoising



Higher-order TV-denoising recovers a better estimate

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}} \|\theta - y\|^2 + \lambda \|D\theta\|_1 \quad \text{— not a linear smoother}$$

◇ Nonparametric Regression on Graphs (d-dim grids)

$$y_i \sim N(\theta_{0,i}, \sigma^2), \quad \text{i.i.d., for } i = 1, \dots, n,$$

- y is observed on every vertex of a graph.
- Estimate θ_0 using noisy observation y.

◇ Optimal rates (d-dim grids, kth order TV)

	$k=0$	$k \geq 1$
$d=1$	$n^{-(2k+2)/(2k+3)}$ (Trend Filter)	
$d>1$	C_n/n (TV-denoising)	??

◇ Questions of interest

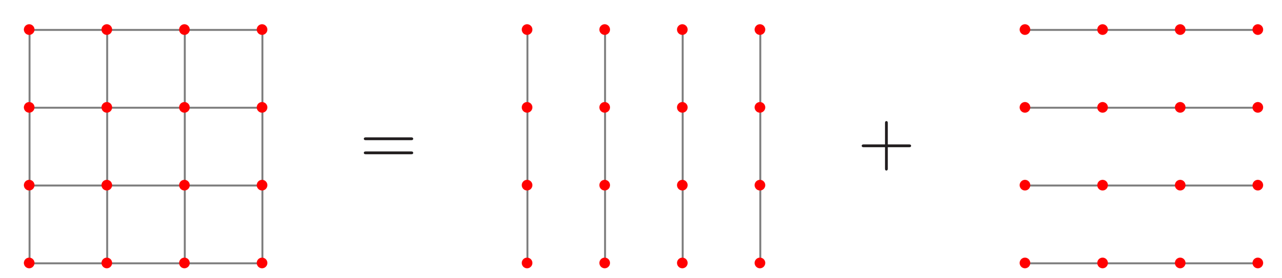
1. What is the discrete analog of kth order TV on grids ($d>1$)?
2. Theoretically quantifying the denoising performance
   - How fast does MSE converge to 0 as we get more pixels?
3. Information-theoretic limit
   - How fast does it get for any method?

Kronecker TF and Graph TF

◇ Kronecker Trend Filtering (KTF)

Penalty is the sum of univariate penalties along rows and columns

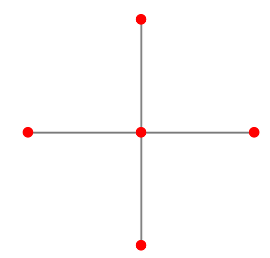
$$\|\Delta_{\mathrm{K}}^{(k+1)}\theta\|_1 = \sum_{j=1}^{N} \|D_{1d}^{(k+1)}\theta_{.j}\|_1 + \sum_{i=1}^{N} \|D_{1d}^{(k+1)}\theta_{i.}\|_1$$



◇ Graph Trend Filtering (GTF) (Wang et al., 2014):

$$\Delta_{\mathrm{G}}^{(1)}, \Delta_{\mathrm{G}}^{(2)}, \Delta_{\mathrm{G}}^{(3)}, \Delta_{\mathrm{G}}^{(4)}, \dots = D, L, DL, L^2, \dots$$

where $L = D^T D$ is the Laplacian of the grid. For $k=1$,

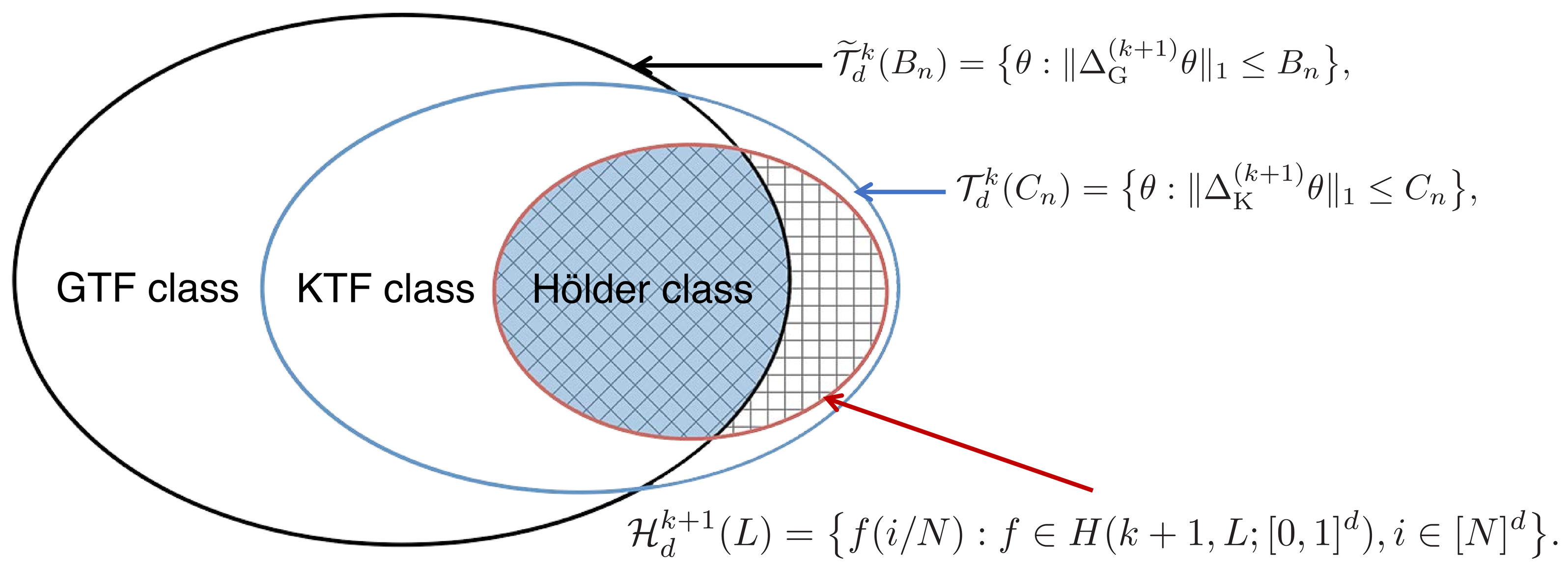


GTF: $|(\theta_1 - 2\theta_0 + \theta_2) + (\theta_3 - 2\theta_0 + \theta_4)|$

KTF: $|\theta_1 - 2\theta_0 + \theta_2| + |\theta_3 - 2\theta_0 + \theta_4|$

- null($\Delta_{\mathrm{K}}^{(k+1)}$): $p \otimes q$ where p, q polynomials of degree $\leq k$
- null($\Delta_{\mathrm{G}}^{(k+1)}$) is $\mathbb{1}$: constant function.

Function classes/Smoothness



- Holder class $\mathcal{H}_d^{k+1}(L) \subseteq$ KTF class $\mathcal{T}_d^k(C_n)$ if $C_n = cn^{1-(k+1)/d}$ (canonical scaling). This delivers a lower bound for KTF class.
- No such embedding for GTF class due to boundary artifacts! Embed an ellipsoid and apply classic results from Donoho, Liu & McGibbon (1990)

Our results

◇ Upper bounds:

$(d=2, k \geq 1)$ if $\|\Delta_{\mathrm{K}}\theta_0\|_1 \leq C_n$, $\|\Delta_{\mathrm{G}}\theta_0\|_1 \leq B_n$

$$\mathrm{MSE}(\widehat{\theta}_K, \theta_0) = \tilde{O}_{\mathbb{P}}\left(\frac{C_n}{n}\right)^{2/(k+2)}, \quad \mathrm{MSE}(\widehat{\theta}_G, \theta_0) = \tilde{O}_{\mathbb{P}}\left(\frac{B_n}{n}\right)^{2/(k+2)}$$

◇ Lower bounds:

For all d, k

$$\mathrm{Risk}(\widetilde{\mathcal{T}}_d(C_n)) = \Omega\Big((C_n/n)^{\frac{2d}{2k+2+d}}\Big), \quad \mathrm{Risk}(\mathcal{T}_d(B_n)) = \Omega\Big((B_n/n)^{\frac{2d}{2k+2+d}}\Big)$$

Matching rates for $d=2, k \geq 1$ up to log factors

◇ Minimax rates under canonical scaling:

	d=1	d=2	d>2
k=0	$n^{-2/3}$	$n^{-2/4}$	$n^{-\frac{1}{d}}$
k=1	$n^{-4/5}$	$n^{-4/6}$	?
k>1	$n^{-\frac{2k+2}{2k+3}}$	$n^{-\frac{2k+2}{2k+4}}$	?

Univariate TF (Tibshirani, 2014) (Mammen& Van De Geer, 2001) → d=1 column

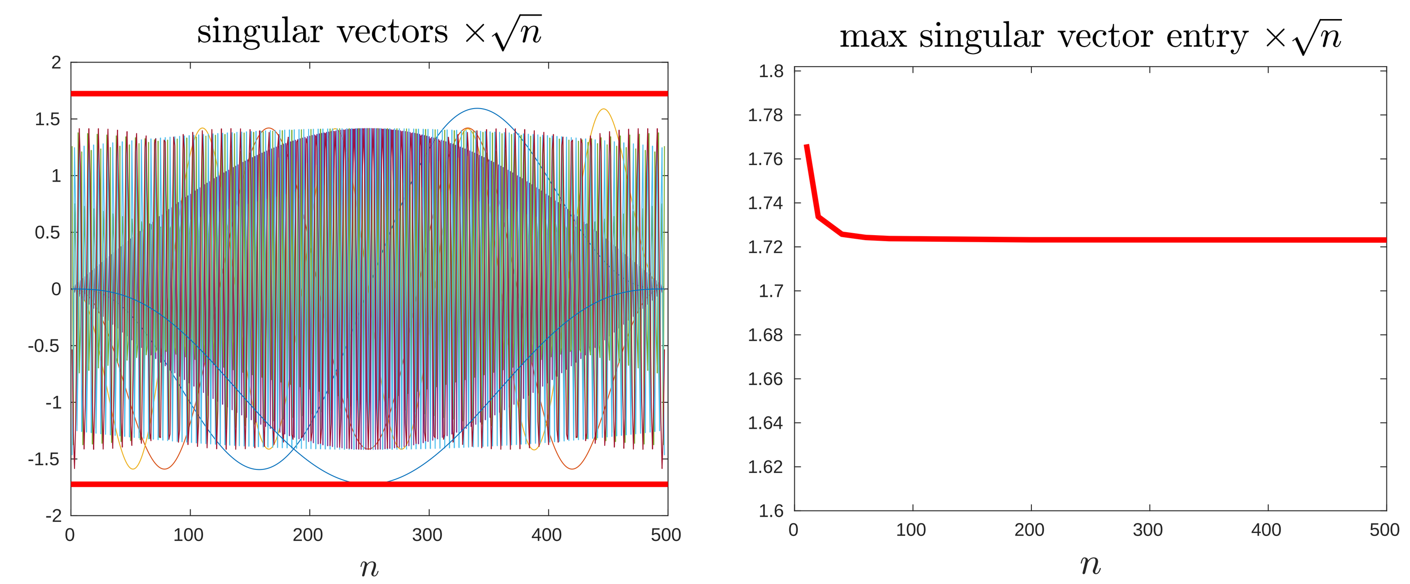
(Sadhanala, Wang, Tibshirani, 2016) → k=0 row, d=2 and d>2

(This paper!) → k=1 and k>1, d=2

Open problem: Minimax rate for d>2, k>1

◇ Upper bound proof ideas:

- Use Theorem 6 of (Wang, Sharpnack, Smola, Tibshirani, 2016).
- $\Delta_{\mathrm{K}}, \Delta_{\mathrm{G}}$ have Kronecker product structure
- Singular vectors are nearly-sinusoidal (challenging to prove!)



- Singular values do not decay too fast

References

[1] Donoho, Liu, and MacGibbon. Minimax Risk Over Hyperrectangles, and Implications *Annals of Statistics*, 18(3):1416–1437, 1990.

[2] Wang, Sharpanack, Smola, Tibshirani. Trend Filtering on Graphs. JMLR, 2016.

[3] Bogoya, Bottcher, Grudsky, and Maximenko. Eigenvectors of Hermitian Toeplitz matrices with smooth simple-loop symbols. *Linear Algebra and its Applications*, 2016.