

## Adaptive rates for trend filtering using dual certificates

Sara van de Geer

Joint work with Francesco Ortelli

The theory for trend filtering has been developed in a series of papers: Tibshirani [2014], Wang et al. [2016], Sadhanala and Tibshirani [2019], Sadhanala et al. [2017], Guntuboyina et al. [2020], and see Tibshirani [2020] for a very good overview and further references. In this course we will combine the approach of Dalalyan et al. [2017] with dual certificates as given in Candes and Plan [2011]. This allows us to obtain new results and extensions.

The trend filtering problem in one dimension is as follows. Let  $Y \sim \mathcal{N}(f^0, I)$  be a vector of observations with unknown mean  $f^0 := \mathbf{E}Y$ . Let for  $f \in \mathbb{R}^n$ ,

$$\begin{aligned} (\Delta f)_i &:= f_i - f_{i-1}, \quad i \geq 2, \\ (\Delta^2 f)_i &:= f_i - 2f_{i-1} + f_{i-2}, \quad i \geq 3, \\ &\vdots \\ (\Delta^k f)_i &:= (\Delta(\Delta^{k-1} f))_i, \quad i \geq k+1. \end{aligned}$$

Then we consider the estimator

$$\hat{f} := \min_{f \in \mathbb{R}^n} \left\{ \|Y - f\|_2^2/n + 2\lambda \|\Delta^k f\|_1 \right\}.$$

Let  $S_0 := \{j : (\Delta^k f^0)_j \neq 0\}$  and  $s_0 := |S_0|$  its size. We want to prove a ‘‘oracle’’ type of result: for an appropriate choice of the tuning parameter, and modulo log-factors, with high probability  $\|\hat{f} - f^0\|_2^2/n \lesssim (s_0 + 1)/n$ . For this one may apply the general theory for the Lasso. Indeed, the above is a Lasso estimator if we write  $f = \Psi b$  where  $\Psi \in \mathbb{R}^{n \times n}$  is the ‘‘design matrix’’ or ‘‘dictionary’’ and where  $b_j = (\Delta^k f)_j$ ,  $j \geq k+1$ . We present the explicit expression for this dictionary and then will notice that the restricted eigenvalue conditions that are typically imposed for Lasso problems do not hold. What we will do instead is use a ‘‘dual certificate’’  $q$  with index set  $\mathcal{D} := \{k+1, \dots, n\}$ . We require that  $q_j = \text{sign}(\Delta^k f^0)_j$  if  $j \in S_0$  and such that  $|q_j| \leq 1 - w_j$  if  $j \in \mathcal{D} \setminus S_0$ , where  $\{w\}_{j \in \mathcal{D} \setminus S_0}$  is a given set of noise weights. Moreover, we require  $q_{k+1} = q_n = 0$ . We call such  $q$  an interpolating vector. We show for an appropriate choice of  $\lambda$

$$\|\hat{f} - f^0\|_2^2/n \lesssim (s_0 + 1)/n + n\lambda^2 \|\Delta^k q\|_2^2$$

and that the two terms on the right hand side can be up to log terms of the same order if the elements in  $S_0$  corresponding to different signs satisfy a minimal distance condition.

We develop a non-asymptotic theory for the problem and refine the above to sharp oracle results. Moreover, the approach we use allows extensions to higher dimensions and to total variation on graphs. For the case of graphs with cycles the main issue is to determine the noise weights, which can be done by counting the number of times an edge is used when traveling from one node to another. Extensions to other loss functions will be considered as well.

## References

- Emmanuel J Candes and Yaniv Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011.
- Arnak Dalalyan, Mohamed Hebiri, and Johannes Lederer. On the prediction performance of the lasso. *Bernoulli*, 23(1):552–581, 2017.
- Adityanand Guntuboyina, Donovan Lieu, Sabayasachi Chatterjee, and Bodhisattva Sen. Adaptive risk bounds in univariate total variation denoising and trend filtering. *Annals of Statistics*, 48(1):205–229, 2020.
- V. Sadhanala and R.J. Tibshirani. Additive models with trend filtering. *Annals of Statistics*, pages 3032–3068, 2019.
- V. Sadhanala, Y.-X. Wang, J. L Sharpnack, and R. J. Tibshirani. Higher-order total variation classes on grids: Minimax theory and trend filtering methods. In *Advances in Neural Information Processing Systems*, pages 5800–5810, 2017.
- Ryan Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. *Annals of Statistics*, 42(1):285–323, 2014.
- Ryan Tibshirani. Divided differences, falling factorials, and discrete splines: Another look at trend filtering and related problems. *arXiv preprint arXiv:2003.03886*, 2020.
- Yu-Xiang Wang, James Sharpnack, Alex Smola, and Ryan Tibshirani. Trend filtering on graphs. *Journal of Machine Learning Research*, 17:15–147, 2016.