Lattice partition recovery with dyadic CART

Oscar Hernan Madrid Padilla*, Yi Yu† and Alessandro Rinaldo‡

*Department of Statistics, University California, Los Angeles
† Department of Statistics, University of Warwick
‡ Department of Statistics & Data Science, Carnegie Mellon University

Motivation

Suppose we observe a noisy realization of a structured piece-wise constant signal supported over a d-dimensional square lattice $L_{d,n} = \{1, \ldots, n\}^d$.

Data that can be modeled in this manner arise in several application areas, especially in surveillance and environment monitoring. Our work is motivated by all the applications/problems considered in the large literature on biclustering. Estimating the boundary of the partition is the most refined and difficult task in these settings.

To be specific, we assume that the data $y \in \mathbb{R}^{L_{d,n}}$ are such that, for each coordinate $i \in L_{d,n}$,

$$y_i = \theta^*_i + \epsilon_i,$$

where $(\epsilon_i, i \in L_{d,n})$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ noise variables and the unknown signal $\theta^* \in \mathbb{R}^{L_{d,n}}$ is assumed to be piece-wise constant over an unknown rectangular partition of $L_{d,n}$.

Rectangular and induced partitions

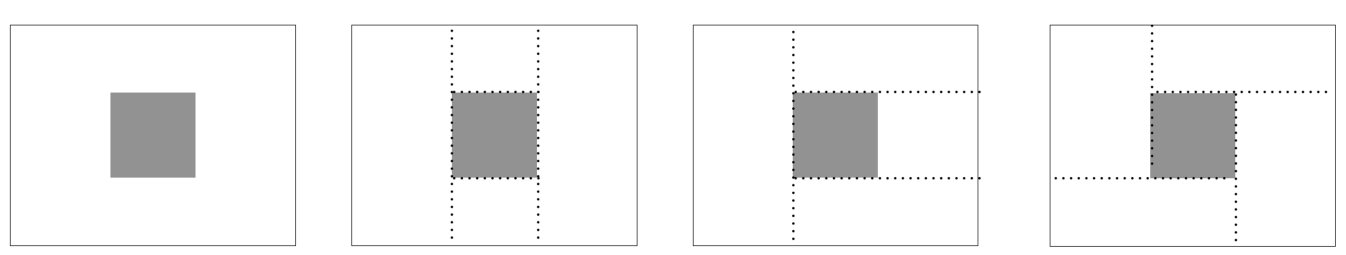
We define a subset $R \subset L_{d,n}$ to be a rectangle if

$$R = \prod_{i=1}^{d} [a_i, b_i],$$

where $[a, b] = \{j \in \mathbb{Z} : a \leq j \leq b\}$.

A rectangular partition of $L_{d,n}$, $\mathcal{P}$, is a collection of disjoint rectangles $\{R_l\} \subset L_{d,n}$, satisfying $\cup_{R \in \mathcal{P}} R = L_{d,n}$.

A rectangular partition associated with a vector $\theta \in \mathbb{R}^{L_{d,n}}$ is a rectangular partition $\{R_l\}_{l \in [1,k]}$ of $L_{d,n}$, such that θ takes on constant values over each R_l. For a vector $\theta \in \mathbb{R}^{L_{d,n}}$, we let $k(\theta)$ be the smallest positive integer such that there exists a rectangular partition with $k(\theta)$ elements and associated with θ.



However, we see from the plots above, rectangular partitions are not necessarily unique. To address this issue, our goal is to recover an induced partition.

Definition (Induced partition) Let $\{R^*_j\}_{j \in [m]}$ be a rectangular partition of $L_{d,n}$ associated with θ^*. Consider the graph $G^* = (E^*, V^*)$, where $V^* = [m]$ and

$$E^* = \{(i, j) : \bar{\theta}^*_{R^*_i} = \bar{\theta}^*_{R^*_j},\ R^*_i \text{ and } R^*_j \text{ are adjacent}\}.$$

Let $\{C^*_l\}_{l \in [L]}$ be all connected components of G^* and define $\Lambda^* = \Lambda^*(\theta^*) = \{\cup_{j \in C^*_1} R^*_j, \ldots, \cup_{j \in C^*_L} R^*_j\}$ as the partition (not necessarily rectangular) induced by θ^*. We say that the union of rectangles $\cup_{j \in C^*_s} R^*_j$ and $\cup_{j \in C^*_t} R^*_j$, $s, t \in [L]$, $s \neq t$, are adjacent, if and only if there exists $(i, j) \in C^*_s \times C^*_t$ such that R^*_i and R^*_j are adjacent.

Let κ and Δ be the minimum jump size and minimal rectangle size,

$$\kappa = \min_{a \in A, b \in B,\ A, B \in \Lambda^*,\ \theta^*_a \neq \theta^*_b} |\theta^*_a - \theta^*_b| \quad \text{and} \quad \Delta = \min_{j \in [m]} |R^*_j|.$$

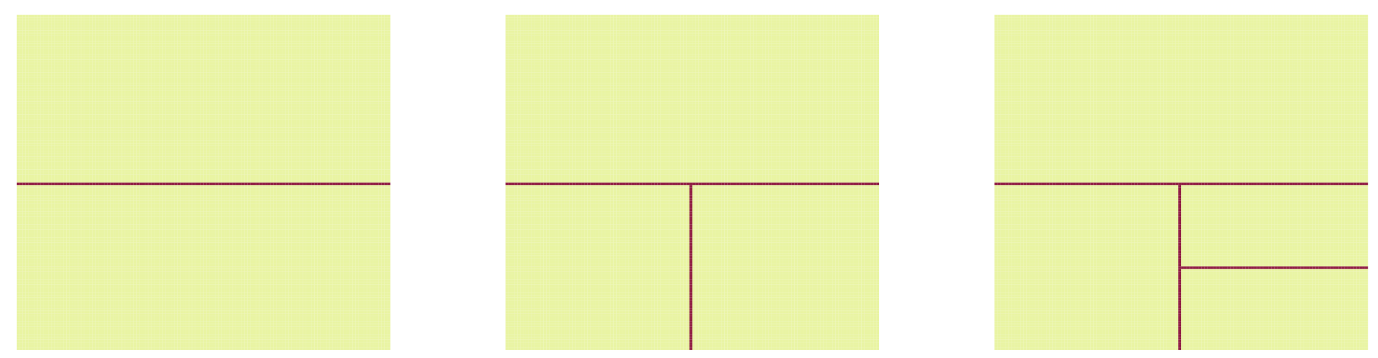
Estimators

Inspired by the success of Potts functional estimators in the change point localizaiton in 1-D lattice, a natural extension would be

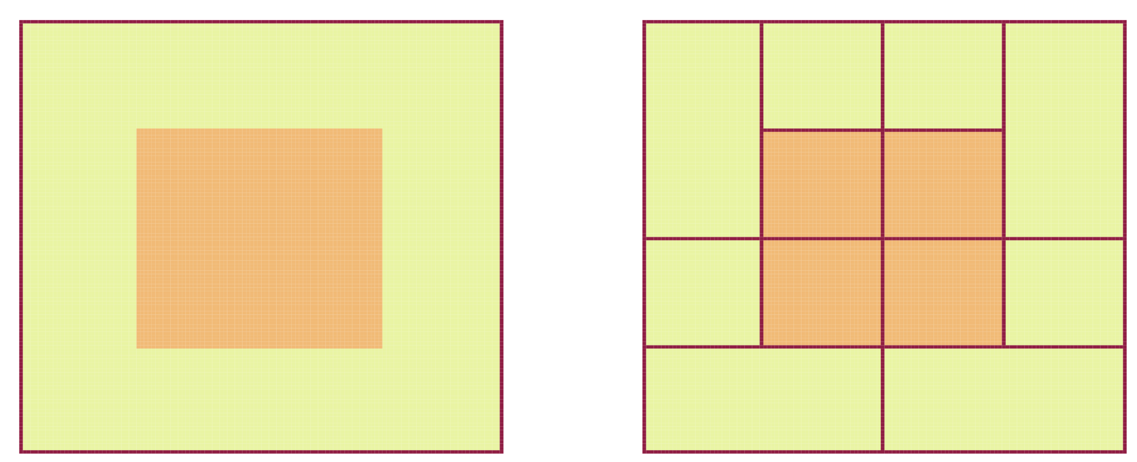
$$\hat{\theta}_1 = \arg\min_{\theta \in \mathbb{R}^{L_{d,n}}} \{2^{-1}\|y - \theta\|^2 + \lambda k(\theta)\},$$

where $\lambda > 0$ is a tuning parameter. Unfortunately, this is an NP-hard estimator when the lattice dimension $d \geq 2$.

Polynomial-time approximations to the above include the optimal regression trees (ORT, 1) and the dyadic classification and regression tree (DCART, 2). In this paper, we focus on DCART. Instead of optimizing over all vectors in $\mathbb{R}^{L_{d,n}}$, DCART minimizes the objective function only over vectors associated with a dyadic rectangular partition, which is illustrated below.



For a vector $\theta \in \mathbb{R}^{L_{d,n}}$, we let $k_{\mathrm{dyadic}}(\theta)$ be the smallest positive integer such that there exists a dyadic rectangular partition with $k_{\mathrm{dyadic}}(\theta)$ elements associated with θ, see below for an example.



One-sided consistency of DCART

Though DCART is known to be minimax optimal in terms of denoising [1], for the partition recovery task, we show that it results in an over-partition, but possesses a one-sided consistency, in the sense that, every resulting DCART rectangle R_i, $i \in [\tilde{m}]$, is almost constant, i.e. there exists $S_i \subset R_i$ such that $\theta^*_t = \theta^*_u$, $u, t \in S_i$ and

$$\sum_{i \in [\tilde{m}]} |R_i \setminus S_i| \lesssim \kappa^{-2}\sigma^2 k_{\mathrm{dyad}}(\theta^*) \log(N).$$

We, however, cannot guarantee that each constant rectangle in signal can be almost covered by an element of DCART estimators.

Two-sided consistency of DCART: A two-step estimator

To address the issue of over-partitioning, we prune the DCART estimator by merging rectangles when their values are similar and the rectangles are not far apart. Let $\widehat{\Lambda}$ be the final output of the induced partition.

Under some regularity conditions, with high probability, it holds that

$$|\widehat{\Lambda}| = |\Lambda^*| \text{ and } d_{\mathrm{Haus}}(\widehat{\Lambda}, \Lambda^*) \lesssim \sigma^2 \kappa^{-2} k_{\mathrm{dyadic}}(\theta^*) \log(N),$$

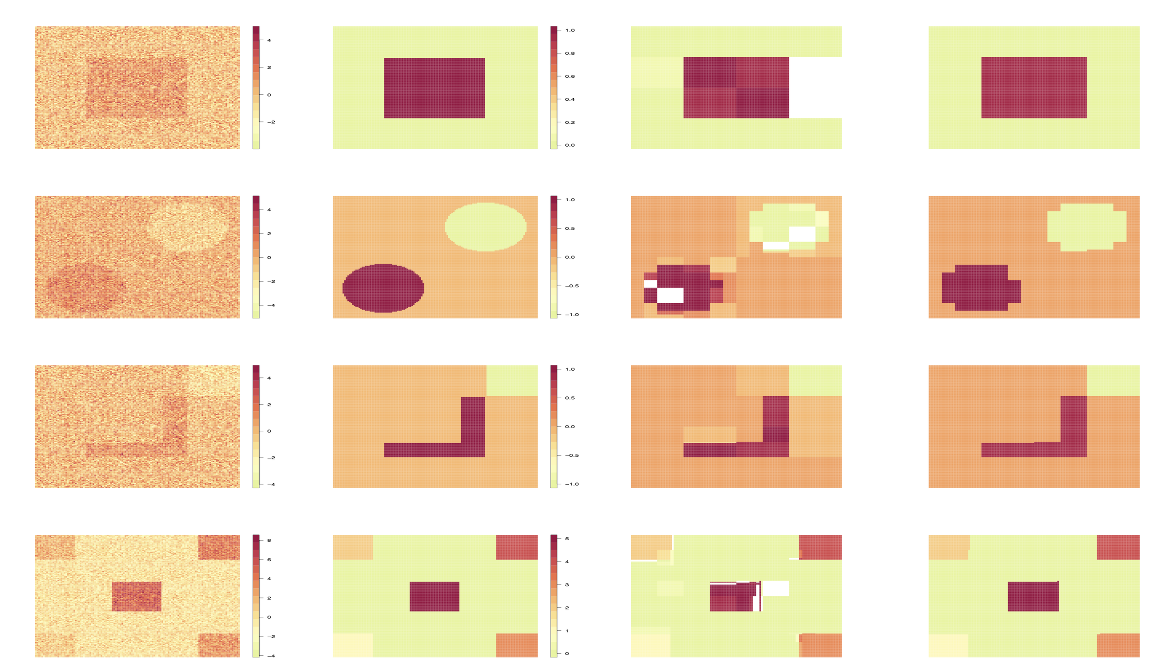
which is consistent in terms of number of elements and in terms of the estimation error rates when $k_{\mathrm{dyadic}}(\theta^*) = O(1)$, supported by our lower bound results.

Optimality: A regular boundary case

We can show, in a specical case, which can be understood as a regular boundary case, that our two-step estimator is minimax optimal in terms of the estimation error rates even when $k_{\mathrm{dyadic}}(\theta^*)$ is allowed to diverge.

This special case can be understood as, each element in Λ^* only has $O(1)$-many elements nearby. It shares the same spirit of requiring the cluster boundary to be a bi-Lipschitz function in the cluster detection literature [e.g. 3].

Simulation results



From left to right: An instance of y, the signal θ^*, DCART and DCART after merging.

Acknowledgements

Funding in direct support of this work: NSF DMS 2015489 and EPSRC EP/V013432/1.

References

[1] Chatterjee, S., & Goswami, S. (2019). Adaptive estimation of multivariate piecewise polynomials and bounded variation functions by optimal decision trees. arXiv preprint arXiv:1911.11562.

[2] Donoho, D. L. (1997). CART and best-ortho-basis: a connection. The Annals of Statistics, 25(5), 1870-1911.

[3] Arias-Castro, E., Candes, E. J., & Durand, A. (2011). Detection of an anomalous cluster in a network. The Annals of Statistics, 278-304.