Guillotine coarseness of bicolored point sets in the plane

José Fernández Goycoolea*[1], Luis H. Herrera†[2], Pablo Pérez-Lantero‡[3], and Carlos Seara§[4]

[1]Departamento de Matemática y Física, Universidad de Magallanes, Punta Arenas, Chile.
[2]Departamento de Informática y Computación, Universidad Tecnológica Metropolitana, Santiago, Chile.
[3]Departamento de Matemática y Ciencia de la Computación, Universidad de Santiago de Chile, Santiago, Chile.
[4]Departament de Matemàtiques, Universitat Politècnica de Catalunya, Barcelona, Spain.

1 Introduction

The concept of the coarseness of a bicolored set of plane points $S = R \cup B$ was originally introduced by Bereg et al. [1] in order to measure how blended the red and blue points are. Let the bichromatic discrepancy $\nabla(X)$ be the absolute difference between red and blue points in any $X \subset S$. Then the coarseness of S is defined to be $\mathcal{C}(S) = \max_{\Pi} \min_{X \in \Pi} \nabla(X)$, where the maximization is taken over the convex partitions Π of S, that is partitions whose members have pairwise nonintersecting convex hulls.

This definition has several desirable properties that relate to the data-mining task of assessing whether a dataset has a tendency to be clustered successfully. However, the direct computation of $\mathcal{C}(S)$ is challenging, even being conjectured as NP-hard, so several modifications have been studied in the literature.

2 Guillotine coarseness

In this context, we present the ideas recently published at Fernandez et al. [3]. A naive, exponential time algorithm is described to compute the general coarseness. Then, two modified definitions are introduced by restricting the available partitions in the maximization of the general definition. The first one, named *Guillotine coarseness* $\mathcal{C}_g(S)$ allows only guillotine partitions of S; i.e., partitions that can be achieved by successive axis-aligned straight cuts of the plane. Then, a further restriction is imposed by allowing only a single step of vertical cuts followed by a step of horizontal cuts or vice versa, this measure is called *Two level guillotine coarseness* $\mathcal{C}_{TLg}(S)$ and we have for any S:

$$\mathcal{C}_{TLg}(S) \leq \mathcal{C}_g(S) \leq \mathcal{C}(S). \tag{1}$$

Leveraging on theoretical properties regarding the coloration of adjacent parts in an optimal guillotine partition and making use of two geometric data structures: range-trees and the MSC-trees of Cortes et al. [2], polynomial time dynamic programming algorithm algorithms are devised for both novel measures, leading to

Theorem 1 *Given a set S of n bicolored points in the plane $\mathcal{C}_g(S)$ can be computed in $O(n^5)$ time and $O(n^4)$ space and $\mathcal{C}_{TLg}(S)$ can be computed in $O(n^2 \log^2 n)$ time and $O(n^2)$ space.*

3 Computational experiments

Finally, computational experiments are presented to demonstrate the performance of the proposed algorithms in practice. The naive approach to $\mathcal{C}(S)$ is implemented and deployed on small instances of the benchmarking dataset proposed by [4]. The dynamic programming approaches for $\mathcal{C}_g(S)$ and $\mathcal{C}_{TLg}(S)$ are tested over the full dataset, showing encouraging results in terms of time performance when compared to the alternative approach of [4].

References

[1] S. Bereg, J. M. Díaz-Báñez, D. Lara, P. Pérez-Lantero, C. Seara, and J. Urrutia. On the coarseness of bicolored point sets. *Computational Geometry: Theory and Applications*, 46(1):65–77, 2013.

[2] C. Cortés, J. M. Díaz-Báñez, P. Pérez-Lantero, C. Seara, J. Urrutia, and I. Ventura. Bichromatic separability with two boxes: A general approach. *Journal of Algorithms*, 64(2):79–88, 2009.

[3] J. Fernández Goycoolea, L. H. Herrera, P. Pérez-Lantero, and C. Seara. Computing the coarseness measure of a bicolored point set over guillotine partitions. *Results in Applied Mathematics*, 24:100503, 2024.

[4] A. Sapucaia, P. J. de Rezende, and C. C. de Souza. Solving the coarseness problem by ILP using column generation. In *International Conference on Computational Science and Its Applications*, pages 30–45. Springer, 2021.

*Email: jose.fernandezg@umag.cl
†Email: luis.herrerab@utem.cl
‡Email: pablo.perez.l@usach.cl Research supported by project DICYT 042332PL, Vicerrectoría de Investigación, Desarrollo e Innovación USACH (Chile).
§Email: carlos.seara@upc.edu Research supported by project PID2019-104129GB-I00 funded by MICIU/AEI, Spain/10.13039/501100011033.