M2 internship at LIGM, Paris Est

Title : Sparse inclusion-exclusion formulas for set-systems with bounded VC dimension

Supervisors :

Xavier Goaoc, Université Paris Est. Email : goaoc@univ-mlv.fr

Nabil Mustafa, ESIEE. Email : mustafan@esiee.fr

The intership would be located at the LIGM laboratory (`ligm.u-pem.fr/`).

Context : The *inclusion-exclusion principle* says that for any family of sets $\{A_1, A_2, \ldots, A_n\}$ we have

$$\mathbb{1}_{\bigcup_{i=1}^n A_i} = \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} \mathbb{1}_{\bigcap_{i\in I} A_i} \tag{1}$$

where $\mathbb{1}_X$ is the indicator function of X. This fundamental property is used to reformulate the computation of volumes, areas, quadrature on unions into similar computations on intersections. It is also the foundation of the *Moebius transform* (also called the *Zeta transform*) which provides the best (exponential-time) algorithm for certain hard optimization problems such as set partitioning and graph coloring [1].

Formula (1) can be drastically simplified in some geometric settings. For instance, if A_i is a family of balls in $\mathbb{R}^d$ and K denotes the *regular triangulation* of the centers weighted by the radii, then

$$\mathbb{1}_{\bigcup_{i=1}^n A_i} = \sum_{\emptyset \neq I \in K} (-1)^{|I|+1} \mathbb{1}_{\bigcap_{i\in I} A_i}. \tag{2}$$

For $d = 2$ this formula has linear size, uses only intersections of up to three sets, and can be computed in $O(n \log n)$ time. To any family of sets $F = \{A_1, A_2, \ldots, A_n\}$, one may associate an hypergraph, called its *Venn diagram*, defined by

$$V(F) = \{I \subseteq [n] \mid (\cap_{i\in I} A_i) \setminus (\cup_{j\in[n]\setminus I} A_j) \neq \emptyset\}$$

Simplified formulas like (2) exist for any family F whose Venn diagram has size less than 2^n [2].

Goals : The goal of this internship is to examine the case where F has bounded VC-dimension. The *VC-dimension* of F is the maximum size of a subset of F with complete Venn diagram :

$$\dim_{\mathrm{VC}}(F) = \max\{|G| : G \subseteq F, |V(G)| = 2^{|G|}\}$$

(this is not the usual formulation, but it is equivalent). Bounded VC-dimension is ubiquitous in computational geometry and is related to many questions in sampling and approximation. It ensures that $V(F)$ has size at most polynomial in n, so sparse formula exist ; the goal is to improve on the general upper bounds of [2] and devise effective algorithms for computing such sparse formulas for this setting.

References

[1] A. Björklund, T. Husfeldt, and M. Koivisto. Set partitioning via inclusion-exclusion. *SIAM J. Comput.*, 39 :546–563, 2009.

[2] X. Goaoc, J. Matoušek, P. Paták, Z.Patákova, and M. Tancer. Simplifying inclusion-exclusion formulas. *Combinatorics, Probability and Computing*, 24 :438–456, 2015.