Labeled sample compression schemes for complexes of oriented matroids

Victor Chepoi[1], Kolja Knauer[1,2], and Manon Philibert[1]

[1]LIS, Aix-Marseille Université, CNRS, and Université de Toulon
Faculté des Sciences de Luminy, F-13288 Marseille Cedex 9, France
{victor.chepoi,kolja.knauer,manon.philibert}@lis-lab.fr

[2]Departament de Matemàtiques i Informàtica, Universitat de Barcelona (UB), Barcelona, Spain

Abstract We show that the topes of a complex of oriented matroids (abbreviated COM) of VC-dimension d admit a proper labeled sample compression scheme of size d. This considerably extends results of Moran and Warmuth on ample classes, of Ben-David and Litman on affine arrangements of hyperplanes, and of the authors on complexes of uniform oriented matroids, and is a step towards the sample compression conjecture – one of the oldest open problems in computational learning theory. On the one hand, our approach exploits the rich combinatorial cell structure of COMs via oriented matroid theory. On the other hand, viewing tope graphs of COMs as partial cubes creates a fruitful link to metric graph theory.

1. Introduction

1.1. **General setting.** Littlestone and Warmuth [50] introduced sample compression schemes as an abstraction of the underlying structure of learning algorithms. Roughly, the aim of a sample compression scheme is to compress samples of a *concept class* $\mathcal{C}$ as much as possible, such that data coherent with the original samples can be reconstructed from the compressed data. There are two types of sample compression schemes: labeled, see [35, 50] and unlabeled, see [7, 34, 48]. A labeled compression scheme of size k compresses every sample of $\mathcal{C}$ to a labeled subsample of size $\leq k$ and an unlabeled compression scheme of size k compresses every sample of $\mathcal{C}$ to a subset of size $\leq k$ of the domain of the sample (see the end of the introduction for precise definitions). The Vapnik-Chervonenkis dimension (*VC-dimension*) of a set system, was introduced by [68] as a complexity measure of set systems. VC-dimension is central in PAC-learning and plays an important role in combinatorics, algorithmics, discrete geometry, and combinatorial optimization. In particular, it coincides with the rank in the theory of (complexes of) oriented matroids. Furthermore, within machine learning and closely tied to the topic of this paper, the *sample compression conjecture* of [35] and [50] states that *any set family of VC-dimension d has a labeled sample compression scheme of size* $O(d)$. This question remains one of the oldest open problems in computational learning theory.

1.2. **Related work.** The best-known general upper bound is due to Moran and Yehudayoff [57] and shows that there exist labeled compression schemes of size $O(2^d)$ for any set family of VC-dimension d. The labeled compression scheme of [57] is not proper and it is even open if there exist proper labeled sample compression schemes which compress samples with support larger than d to subsamples with strictly smaller support [55]. From below, Floyd and Warmuth [35] showed that there are classes of VC-dimension d admitting no labeled compression scheme of size less than d and that no concept class of VC-dimension d admits a labeled compression scheme of size at most $\frac{d}{5}$. Pálvölgyi and Tardos [63] exhibited a concept class of VC-dimension 2 with no unlabeled compression scheme of size 2. However, no similar results are known for labeled sample compression schemes. Prior to [63], it was shown in [60] that the concept class of positive halfspaces in $\mathbb{R}^2$ (which has VC-dimension 2) does not admit proper unlabeled sample compression schemes of size 2.

For more structured concept classes better upper bounds are known. Ben-David and Litman [7] proved a compactness lemma, which reduces the existence of labeled or unlabeled compression schemes for arbitrary concept classes to finite concept classes. They also obtained unlabeled compression schemes for regions in arrangements of affine hyperplanes (which correspond to realizable Affine Oriented Matroids in our language). Finally, they obtained sample compression schemes for concept classes by embedding them into concept classes for which such schemes were known. Helmbold, Sloan, and Warmuth [43] constructed unlabeled compression schemes of size d for intersection-closed concept classes of VC-dimension d. They compress each sample to a minimal generating set and show that the size of this set is upper bounded by the VC-dimension. An important class for which positive results are available is given by ample set systems [3, 27] (originally introduced as lopsided sets by Lawrence [49]). They capture an important variety of combinatorial objects, e.g., (conditional) antimatroids, see [29], diagrams of (upper locally) distributive lattices, median graphs or CAT(0) cube complexes, see [3] and were rediscovered in various disguises, e.g. by [10] as *extremal for (reverse) Sauer* and by [58] as *shattering-extremal* [58]. Moran and Warmuth [56] provide labeled sample compression schemes of size d for ample set systems of VC-dimension d. For maximum concept classes (a subclass of ample set systems) unlabeled sample compression schemes of size d have been designed by Chalopin et al. [11]. They also characterized unlabeled compression schemes for ample classes via the existence of *unique sink orientations* of their graphs. However, the existence of such orientations remains open.

1.3. **OMs and COMs.** A structure somewhat opposed to ample classes are Oriented Matroids (OMs), see the book of Björner et al. [8]. Co-invented by Bland and Las Vergnas [9] and Folkman and Lawrence [36], and further investigated by Edmonds and Mandel [30] and many other authors, oriented matroids represent a unified combinatorial theory of orientations of ordinary matroids, which simultaneously captures the basic properties of sign vectors representing the regions in a hyperplane arrangement in $\mathbb{R}^d$ and of sign vectors of the circuits in a directed graph. OMs provide a framework for the analysis of combinatorial properties of geometric configurations occurring in discrete geometry and in machine learning. Point and vector configurations, order types, hyperplane and pseudo-line arrangements, convex polytopes, directed graphs, and linear programming find a common generalization in this language. The Topological Representation Theorem of [36] connects the theory of OMs on a deep level to arrangements of pseudohyperplanes and distinguishes it from the theory of ordinary matroids.

Complexes of Oriented Matroids (COMs) were introduced by Bandelt, Chepoi, and Knauer [4] as a natural common generalization of ample classes and OMs. Ample classes are exactly the COMs with cubical cells, while OMs are the COMs with a single cell. In general COMs, the cells are OMs and the resulting cell complex is contractible. In the realizable setting, a COM corresponds to the intersection pattern of a hyperplane arrangement with an open convex set, see Figure 1. Examples of COMs neither contained in the class of OMs nor in ample classes include linear extensions of a poset or acyclic orientations of mixed graphs, see [4], CAT(0) Coxeter complexes of [40], hypercellular and Pasch graphs, see [17], and Affine Oriented Matroids through [6] and [23]. Note that none of the listed examples is contained in the classes of OMs or ample classes. Apart from the above, COMs have spiked research and appear as the next natural class to attack in different areas such as combinatorial semigroup theory by [53], algebraic combinatorics in relation to the Varchenko determinant by [44], neural codes [45], poset cones see [26], as well as sweeping sequences see [62]. In particular, relations to COMs have already been established within sample compression, by [18, 19, 52] and [11]. A central feature of COMs is that they can be studied via their tope graphs, see Figure 1 for an example. Indeed, the recent characterization of their tope graphs by [46] establishes an embedding of the theory of COMs into metric graph theory, with theoretical and algorithmic implications. Namely, tope graphs of COMs form a subclass of the ubiquitous metric graph class of partial cubes, i.e., isometric subgraphs of hypercubes, with applications ranging from interconnection networks [38] over media theory [33], to chemical graph theory [32]. On the other hand, tope graphs of COMs can be recognized in polynomial time [31, 46]. The graph theoretic view has been used in several recent publications, see [16, 47, 51] and is essential to the current work.

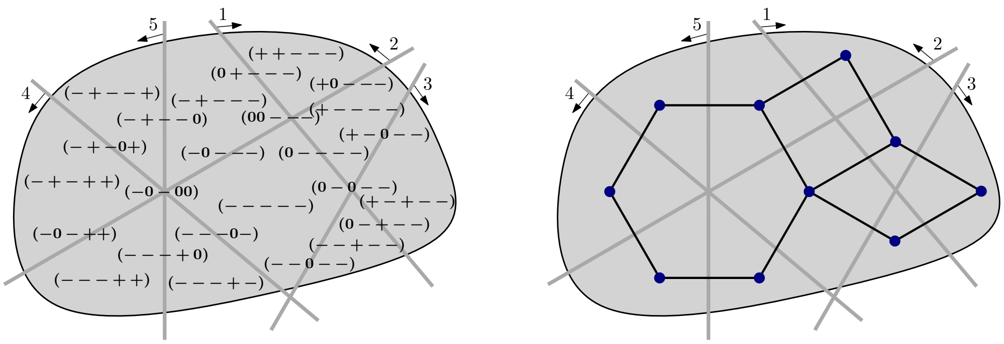

FIGURE 1. A realizable COM and its tope graph.

1.4. **Labeled sample compression schemes.** As we explain later, COMs can be defined as sets of sign vectors, which is another unifying feature for OMs and ample classes. This turns out to be beneficial for the present paper, since the language of sign vectors is perfectly suited for defining sample compression schemes formally. The following formulation is due to [12], for classical formulations, see [50, 56, 57]. Let U be a finite set, called the *universe* and $\mathcal{C}$ be a family of subsets of U, called a *concept class*. We view $\mathcal{C}$ as a set of $\{-1,+1\}$-vectors, i.e., $\mathcal{C} \subseteq \{-1,+1\}^U$. We also consider sets of $\{-1,0,+1\}$-vectors, i.e., subsets of $\{-1,0,+1\}^U$ endowed with the product order $\leq$ between sign vectors relative to the ordering $0 \leq -1,+1$. The sign vectors of the set $\downarrow\mathcal{C} = \bigcup_{C \in \mathcal{C}} \{S \in \{-1,0,+1\}^U : S \leq C\}$ are *realizable samples* for $\mathcal{C}$. A *labeled sample compression scheme* of size k for a concept class $\mathcal{C} \subseteq \{-1,+1\}^U$ is a pair (α,β) of mappings, where $\alpha : \downarrow\mathcal{C} \to \{-1,0,+1\}^U$ is called *compression function* and $\beta : \{-1,0,+1\}^U \to \{-1,+1\}^U$ the *reconstruction function* such that for any realizable sample $S \in \downarrow\mathcal{C}$, it holds $\alpha(S) \leq S \leq \beta(\alpha(S))$ and $|\underline{\alpha}(S)| \leq k$, where $\underline{\alpha}(S)$ is the support of the sign vector $\alpha(S)$, i.e., the non-zero entries of $\alpha(S)$. The condition $S \leq \beta(\alpha(S))$ means that the restriction of $\beta(\alpha(S))$ on the support of S coincides with the input sample S. In particular, if S is a concept of $\mathcal{C}$, then $\beta(\alpha(S)) = S$, i.e., the reconstructor must reconstruct the input concept. A labeled sample compression scheme is *proper* if $\beta(\alpha(S)) \in \mathcal{C}$ for all $S \in \downarrow\mathcal{C}$. Notice that the labeled compression schemes of size $O(2^d)$ of [57] are not proper (i.e., $\beta(\alpha(S))$ is not necessarily a concept of $\mathcal{C}$) and they use additional information. The compression schemes developed in [12] for balls in graphs are proper but also use additional information. The *unlabeled sample compression schemes* [48] (which are not the subject of this paper) are defined analogously, with the difference that in the unlabeled case $\alpha(S)$ is a subset of size at most k of the support of S.

The definition of labeled compression scheme implies that if $\mathcal{C}' \subseteq \mathcal{C}$ and (α,β) is a labeled sample compression scheme for $\mathcal{C}$, then (α,β) is a labeled sample compression scheme for $\mathcal{C}'$. However, (α,β) is in general not proper for $\mathcal{C}'$. Still, this yields an approach (suggested in [66] and implicit in [35]) to obtain improper schemes. For instance, using the result of [56] that ample classes of VC-dimension d admit labeled sample compression schemes of size d, one can try to extend a given set system to an ample class without increasing the VC-dimension too much and then apply their result. In [18] it is shown that partial cubes of VC-dimension 2 can be extended to ample classes of VC-dimension 2. Furthermore, in [19] it is shown that OMs and complexes of uniform oriented matroids (CUOMs) can be extended to ample classes without increasing the VC-dimension. Thus, in these classes there exist improper labeled sample compression schemes whose size is the VC-dimension. On the other hand, there exist partial cubes of VC-dimension 3 that cannot be extended to ample classes of VC-dimension 3, see [19], as well as set systems of VC-dimension 2, that cannot be extended to partial cubes of VC-dimension 2, see [18]. In [19] it is conjectured that every COM of VC-dimension d can be extended to an ample class of VC-dimension d. This would yield improper labeled sample compression schemes for COMs of size d.

1.5. **Our result.** In this paper, we follow a different strategy to give (stronger) proper labeled sample compression schemes of size d for general COMs of VC-dimension d, see Theorem 3. More precisely, we show that the set systems defined by the topes of COMs satisfy the strong form of the sample compression conjecture, i.e., COMs of VC-dimension d admit *proper labeled sample compression schemes of size d.*

Our work substantially extends the result of [56] for ample concept classes, the result of [7] for concept classes arising from arrangements of affine hyperplanes (i.e., realizable Affine Oriented Matroids), and our results [19] for OMs and CUOMs. Many classes of COMs are only covered by this new result. For example, the classes of COMs mentioned in Subsection 1.3 are neither ample, nor affine, nor uniform. Some of these examples are realizable and can be embedded into realizable Affine Oriented Matroid to which one can apply the result of [7]. However, this will lead only improper compression schemes. One important class of COMs, which is neither realizable, nor ample, nor affine, nor uniform, is the class of non-realizable OMs. By the *Topological Representation Theorem of Oriented Matroids* of Folkman and Lawrence [36], the topes of OMs can be characterized as the inclusion maximal cells of an arrangement of pseudohyperplanes. An OM is non-realizable if it it represented by a non-stretchable arrangement, i.e., an arrangement whose pseudohyperplanes cannot be replaced by linear hyperplanes.

To illustrate the representation by pseudohyperplanes, in Figure 2 we give an example of an arrangement U of pseudolines in $\mathbb{R}^2$ and its graph of regions, i.e., the tope graph of the resulting COM. While this example is stretchable, there are many non-stretchable arrangements. Indeed, most OMs are non-realizable [8, Theorems 7.4.2 and 8.7.5]. Deciding stretchability of a pseudoline arrangement and more generally realizability of an OM is a complete problem of the existential theory of the reals, hence in particular NP-hard, see [67]. By a result of Edmonds and Mandel [30], all arrangements of pseudohyperplanes can be considered piecewise-linear.

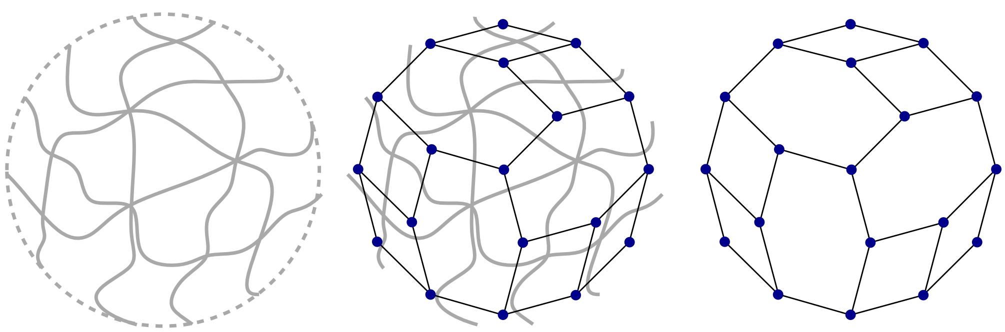

FIGURE 2. A pseudoline arrangement U and its region graph.

1.6. **Pseudohyperplane arrangements and Machine Learning.** Pseudohyperplane arrangements have already arisen in the context of sample compression schemes and VC-dimension in [37, 54, 64, 65] in the treatment of maximum and ample classes. More recently, particular piecewise-linear pseudohyperplane arrangements and their regions occurred in the study of deep feedforward neural networks with ReLU activations [24, 39, 41, 42, 59]. In this theory they appear under the names "arrangements of bent hyperplanes" and "activation regions", respectively. Recall that a (trained) feedforward neural network used to answer Yes/No (i.e., $\{-1, +1\}$) classification problems is a particular type of function $F : \mathbb{R}^d \to \mathbb{R}$. The inputs to F are data feature vectors and the outputs are used to answer the binary classification problem by partitioning the input space $\mathbb{R}^d$ into activation regions.

Next, we closely follow [39] and [42]. A ReLU function $\mathrm{ReLU} : \mathbb{R} \to \mathbb{R}$ is defined by $\mathrm{ReLU}(x) = \max\{0, x\}$. ReLU is among the most popular activation functions for deep neural networks. Let $\sigma : \mathbb{R}^d \to \mathbb{R}^d$ denote the function that applies ReLU to each coordinate. Let $n_0, \ldots, n_k, n_{k+1} = 1$ be a sequence of natural numbers and let $A_i : \mathbb{R}^{n_{i-1}} \to \mathbb{R}^{n_i}, i = 1, \ldots, k+1$ be (parametrized) affine maps. A ReLU (Rectified Linear Unit) network $\mathcal{N}$ of architecture $(n_0, \ldots, n_k)$, depth $k+1$, and $n := \sum_{i=0}^{m} n_i$ neurons is a neural network in which the map F is defined as the composition of the layer maps $F_1 = \sigma \circ A_1, \ldots, F_k = \sigma \circ A_k, F_{k+1} = A_{k+1}$. An activation pattern for $\mathcal{N}$ is an assignment of a $\{-1, +1\}$-sign to each neuron. Given a vector θ of trainable parameters, the activation pattern of the neurons defines a partition of the input space $\mathbb{R}^d$ into activation regions. The activation regions can be viewed as the regions defined by the arrangement of bent hyperplanes associated to layers; for the precise definition see [39, Section 6] and [42]. Activation regions are convex polyhedra [42] and one of important questions in the complexity analysis of deep ReLU networks is counting the number of such activation regions [41, 42, 59]. Notice that the arrangements of bent hyperplanes may not be arrangements of pseudohyperplanes in the classical sense [8] because two bent hyperplanes may not intersect transversally. Transversality of arrangements of bent hyperplanes was investigated in depth in the recent paper [39]. It will be interesting to further investigate how sample compression schemes can be useful in the setting of deep ReLU networks.

2. Preliminaries

2.1. **OMs and COMs.** We recall the basic theory OMs and COMs from [8] and [4], respectively. Let U be a set of size m and let $\mathcal{L}$ be a *system of sign vectors*, i.e., maps from U to $\{-1, 0, +1\}$. The elements of $\mathcal{L}$ are referred to as *covectors* and denoted by capital letters X, Y, Z. For $X \in \mathcal{L}$, the subset $\underline{X} = \{e \in U : X_e \neq 0\}$ is the *support* of X and its complement $X^0 = U \setminus \underline{X} = \{e \in U : X_e = 0\}$ is the *zero set* of X. For a sign vector X and a subset $A \subseteq U$, let X_A be the restriction of X to A. For $X, Y \in \mathcal{L}$, $\mathrm{Sep}(X, Y) = \{e \in U : X_e Y_e = -1\}$ is the *separator* of X and Y. The *composition* of X and Y is the sign vector $X \circ Y$, where for all $e \in U$, $(X \circ Y)_e = X_e$ if $X_e \neq 0$ and $(X \circ Y)_e = Y_e$ if $X_e = 0$.

Definition 1. An *oriented matroid* (OM) is a system of sign vectors $\mathcal{M} = (U, \mathcal{L})$ satisfying

- **(C)** (Composition) $X \circ Y \in \mathcal{L}$ for all $X, Y \in \mathcal{L}$.
- **(SE)** (Strong elimination) for each pair $X, Y \in \mathcal{L}$ and for each $e \in \mathrm{Sep}(X, Y)$, there exists $Z \in \mathcal{L}$ such that $Z_e = 0$ and $Z_f = (X \circ Y)_f$ for all $f \in U \setminus \mathrm{Sep}(X, Y)$.
- **(Sym)** (Symmetry) $-\mathcal{L} = \{-X : X \in \mathcal{L}\} = \mathcal{L}$, that is, $\mathcal{L}$ is closed under sign reversal.

We only consider *simple* systems of sign-vectors $\mathcal{L}$, i.e., if for each $e \in U$, $\{X_e : X \in \mathcal{L}\} = \{-1, 0, +1\}$ and for all $e \neq f$ there exist $X, Y \in \mathcal{L}$ with $\{X_e X_f, Y_e Y_f\} = \{+1, -1\}$.

Let $\leq$ be the product ordering on $\{-1, 0, +1\}^U$ relative to the ordering $0 \leq -1, +1$. The poset $(\mathcal{L}, \leq)$ of an OM $\mathcal{M}$ with an artificial global maximum $\hat{1}$ forms the (graded) *big face lattice* $\mathcal{F}_{\mathrm{big}}(\mathcal{M})$. The length of maximal chains of $\mathcal{F}_{\mathrm{big}}(\mathcal{M})$ minus 1 is the *rank* of $\mathcal{L}$ and denoted $\mathrm{rank}(\mathcal{M})$. The rank of the underlying matroid $\underline{\mathcal{M}}$ equals $\mathrm{rank}(\mathcal{M})$ [8, Thm 4.1.14].

The *topes* $\mathcal{T}$ of $\mathcal{M}$ are the co-atoms of $\mathcal{F}_{\mathrm{big}}(\mathcal{M})$. By simplicity the topes are $\{-1, +1\}$-vectors and $\mathcal{T}$ can be seen as a family of subsets of U. For each $T \in \mathcal{T}$, an element $e \in U$ belongs to the corresponding set if and only if $T_e = +1$. The *tope graph* $G(\mathcal{M})$ of an OM $\mathcal{M}$ is the 1-inclusion graph of the set $\mathcal{T}$ of topes of $\mathcal{L}$, i.e., the subgraph of the hypercube $\mathrm{Q}(U)$ induced by the vertices corresponding to $\mathcal{T}$, see Figure 1.

In *realizable OMs* (i.e., of OMs arising from central hyperplane arrangements of $\mathbb{R}^d$), $X \leq Y$ for two covectors X, Y if and only if the cell corresponding to X is contained in the cell corresponding to Y. Consequently, the topes of realizable OMs are the covectors of the inclusion maximal cells (which all have dimension d), called *regions*. Therefore, the tope graph of a realizable OM can be viewed as the adjacency graph of regions: the vertices of this graph are the regions of a hyperplane arrangement and two regions are adjacent in this graph if they are separated by a unique hyperplane of the arrangement. The *Topological Representation Theorem of Oriented Matroids* of [36], generalizes this correspondence to all OMs: tope graphs of OMs can be characterized as the adjacency graphs of maximal cells of pseudohyperplane arrangements in $\mathbb{R}^d$ [8], where d is the rank of the OM. More precisely, two topes are adjacent if and only if the corresponding regions are separated by a unique pseudohyperplane, see Figure 1. It is also well-known (see for example [8]) that $\mathcal{L}$ can be recovered from its tope graph $G(\mathcal{L})$ (up to isomorphism). Therefore, *we can define all terms in the language of tope graphs.*

Another important axiomatization of OMs is in terms of *cocircuits* of $\mathcal{L}$. These are the atoms of $\mathcal{F}_{\text{big}}(\mathcal{L})$. Their collection is denoted by $\mathcal{C}^*$ and axiomatized as follows: a system of sign vectors $(U, \mathcal{C}^*)$ is an *oriented matroid* (OM) if $\mathcal{C}^*$ satisfies (Sym) and the two axioms:

- **(Inc)** (Incomparability) $\underline{X} \subseteq \underline{Y}$ implies $X = \pm Y$ for all $X, Y \in \mathcal{C}^*$.
- **(E)** (Elimination) for each pair $X, Y \in \mathcal{C}^*$ with $X \neq -Y$ and for each $e \in \mathrm{Sep}(X, Y)$, there exists $Z \in \mathcal{C}^*$ such that $Z_e = 0$ and $Z_f \in \{0, X_f, Y_f\}$ for all $f \in U$.

The set $\mathcal{L}$ of covectors can be derived from $\mathcal{C}^*$ by taking the closure of $\mathcal{C}^*$ under composition.

COMs are defined by replacing the global axiom (Sym) with a weaker local axiom:

Definition 2. A *complex of oriented matroids* (COMs) is a system of sign vectors $\mathcal{M} = (U, \mathcal{L})$ satisfying (SE) and the following axiom:

- **(FS)** (Face symmetry) $X \circ -Y \in \mathcal{L}$ for all $X, Y \in \mathcal{L}$.

One can see that OMs are exactly the COMs containing the zero vector $\mathbf{0}$, see [4]. The twist between (Sym) and (FS) allows to keep on using the same concepts, such as topes, tope graphs, the sign-order and the big face (semi)lattice in a completely analogous way. On the other hand, it leads to a combinatorial and geometric structure that is build from OMs as cells but is much richer than OMs. Let $\mathcal{M} = (U, \mathcal{L})$ be a COM and $X \in \mathcal{L}$ a covector. The *face* of X is $\uparrow X := \{Y \in \mathcal{L} : X \leq Y\}$, sometimes denoted $\mathrm{F}(X)$, see [4,8]. A *facet* is an inclusion maximal proper face. From the definition, any face $\uparrow X$ consists of the sign vectors of all faces of the subcube of $[-1, +1]^U$ with barycenter X. By [4, Lemma 4], each face $\uparrow X$ of a COM $\mathcal{M}$ is an OM, which we denote $\mathcal{M}(X)$. *Ample classes* (called also lopsided [3,49] or extremal [10,56]) are exactly the COMs, in which all faces are cubes. Since OMs are COMs, each face of an OM is an OM and the facets correspond to cocircuits. Furthermore, by [4, Section 11] replacing each combinatorial face $\uparrow X$ of $\mathcal{L}$ by a PL-ball, we obtain a contractible cell complex associated to each COM. The *topes* $\mathcal{T}$ and the *tope graph* $G_{\mathcal{M}}$ of a COM are defined as for OMs. Again, the COM $\mathcal{M}$ can be recovered from $G(\mathcal{M})$, see [4,46]. For $X \in \mathcal{L}$, the topes in $\uparrow X$ induce a subgraph $[X]$ of $G(\mathcal{M})$, which is isomorphic to the tope graph of $\mathcal{M}(X)$ and it is crucial for this paper.

2.2. **Realizable COMs.** In this subsection, we recall the geometric illustration of the axioms in the case of realizable COMs given in the paper [4]. Let U be an affine arrangement of hyperplanes of $\mathbb{R}^d$ and C an open convex set. Restrict the arrangement pattern to C, that is, remove all sign vectors which represent the open regions disjoint from C. Denote the resulting set of sign vectors by $\mathcal{L}(U, C)$ and call it a *realizable COM*. If U is a central arrangement with C being any open convex set containing the origin, then $\mathcal{L}(E, C)$ coincides with the realizable oriented matroid of U. If the arrangement U is affine and C is the entire space, then $\mathcal{L}(U, C)$ coincides with the realizable affine oriented matroid of U. The realizable ample sets arise by taking the central arrangement U of all coordinate hyperplanes U restricted to an arbitrary open convex set C of $\mathbb{R}^d$ (this model was first considered in [49]).

We argue, why a realizable COM satisfies the axioms from Definition 2. Let X and Y be sign vectors belonging to $\mathcal{L}(U, C)$ and designating two regions of C defined by U. Let x, y be two points in these regions. Connect x, y by a line segment and choose $\epsilon > 0$ so that the open ball of radius ϵ around x is contained in C and intersects only those hyperplanes from U containing x. Pick any point w from the intersection of this ϵ-ball with the open line segment between x and y. The corresponding sign vector W is the composition $X \circ Y$, establishing (C). If we select instead a point v on the ray from y via x within the ϵ-ball but beyond x, then the corresponding sign vector V has the opposite signs as W at the coordinates corresponding to the hyperplanes from U containing x and not including the ray from y via x. Hence, $V = X \circ -Y$, yielding the axiom (FS). Now, assume that the hyperplane e from U separates x and y, that is, the line segment between x and y crosses e at some point z. The corresponding sign vector Z is then zero at e and equals the composition $X \circ Y$ at all coordinates where X and Y are sign-consistent, establishing the axiom (SE). If the hyperplanes of U have a non-empty intersection, then any point o from this intersection corresponds to the zero sign vector, showing that central hyperplane arrangements define OMs. In this case, $\mathcal{L}(U, C)$ coincides with $\mathcal{L}(U, \mathbb{R}^d)$ as well as with $\mathcal{L}(U, C_\epsilon)$, where C_ϵ is any open ball centered at o. The face $\uparrow X$ of a covector $X \in \mathcal{L}(U, C)$ is obtained by taking any point $x \in C$ corresponding to X and a small ϵ-ball C_ϵ centered at x. Then $\uparrow X$ coincides with the OM $\mathcal{L}(U, C_\epsilon)$. Finally, notice that the topes of $\mathcal{L}(U, C)$ correspond to the connected components of C minus the hyperplanes of U. Two such topes are adjacent in the tope graph if and only if the corresponding regions are separated by a single hyperplane. That tope graphs of realizable COMs $\mathcal{L}(U, C)$ with $C = \mathbb{R}^d$ was proved by Deligne [21, Proposition 1.3].

2.3. **Deletion and duality.** We continue with restrictions and deletions in OMs and COMs. Let $\mathcal{M} = (U, \mathcal{L})$ be a COM and $A \subseteq U$. Given a sign vector $X \in \{\pm 1, 0\}^U$ by $X \setminus A$ (or by $X_{|U \setminus A}$) we refer to the *restriction* of X to $U \setminus A$, that is $X \setminus A \in \{\pm 1, 0\}^{U \setminus A}$ with $(X \setminus A)_e = X_e$ for all $e \in U \setminus A$. The *deletion* of A is defined as $\mathcal{M} \setminus A = (U \setminus A, \mathcal{L} \setminus A)$, where $\mathcal{L} \setminus A := \{X \setminus A : X \in \mathcal{L}\}$. We often consider the following type of deletion. For a covector $X \in \mathcal{L}$, we denote by $\mathcal{M}(X) = (U \setminus \underline{X}, \uparrow X \setminus \underline{X})$ the OM defined by the face $\uparrow X$, i.e., $\mathcal{M}(X) = \mathcal{M} \setminus \underline{X}$. The classes of COMs and OMs are closed under deletion, see [4, Lemma 1]. The cocircuits and the covectors of deletions of OMs are described in the following way:

Lemma 1. *[8] Let $\mathcal{M} = (U, \mathcal{L})$ be an* OM *with the set of cocircuits $\mathcal{C}^*$ and $A \subseteq U$. Then the cocircuits of $\mathcal{M} \setminus A$ are $\mathcal{C}^* \setminus A$ and the covectors of $\mathcal{M} \setminus A$ are $\mathcal{L} \setminus A$.*

We briefly recall the duality of OMs, see [8, Section 3.4]. The duality is defined via orthogonality of circuits and cocircuits, which can be viewed as a synthetic version of classical orthogonality of vectors. Two sign-vectors $X, Y \in \{\pm 1, 0\}^U$ are *orthogonal*, denoted $X \perp Y$, if either $\underline{X} \cap \underline{Y} = \varnothing$ or there are $e, f \in \underline{X} \cap \underline{Y}$ such that $X_e Y_e = -X_f Y_f$. Oriented matroids can be defined in terms of their *vectors* $\mathcal{V}$ and *circuits* $\mathcal{C}$, which can be derived from the cocircuits $\mathcal{C}^*$ using the following result:

Theorem 1. *[8, Theorem 3.4.3 and Proposition 3.7.12] The set $\mathcal{V}$ consists of all $Y \in \{\pm 1, 0\}^U \setminus \{\boldsymbol{0}\}$ such that $Y \perp X$ for any $X \in \mathcal{C}^*$ and $\mathcal{C}$ consists of the minimal members of $\mathcal{V}$.*

2.4. **Partial cubes and pc-minors.** It is well-known see for example [4, 8] that tope graphs of OMs and COMs are partial cubes, which we introduce now. Let $G = (V, E)$ be a finite, connected, simple graph. The *distance* $d(u, v) := d_G(u, v)$ between vertices u and v is the length of a shortest (u, v)-path, and the *interval* $I(u, v) := \{x \in V : d(u, x) + d(x, v) = d(u, v)\}$ consists of all vertices on shortest (u, v)-paths. A subgraph H is *convex* if $I(u, v) \subseteq H$ for any $u, v \in H$ and *gated* [28] if for every vertex $x \notin H$ there exists a vertex x' (the *gate* of x) in H such that $x' \in I(x, y)$ for each vertex y of H. It is easy to see that gates are unique and that gated sets are convex. An induced subgraph H of G is *isometric* if the distance between vertices in H is the same as that in G. A graph $G = (V, E)$ is *isometrically embeddable* into a graph $H = (W, F)$ if there exists $\varphi : V \to W$ such that $d_H(\varphi(u), \varphi(v)) = d_G(u, v)$ for all $u, v \in V$. A graph G is a *partial cube* if it admits an isometric embedding into a hypercube Q_m. For an edge uv of G, let $W(u,v) = \{x \in V : d(x,u) < d(x,v)\}$. For an edge uv, the sets $W(u,v)$ and $W(v,u)$ are called *complementary halfspaces* of G.

Theorem 2. *[25] A graph G is a partial cube if and only if G is bipartite and for any edge uv the sets $W(u,v)$ and $W(v,u)$ are convex.*

Djoković [25] introduced the following binary relation Θ on the edges of G: for two edges $e = uv$ and $e' = u'v'$, we set $e\Theta e'$ if $u' \in W(u,v)$ and $v' \in W(v,u)$. If G is a partial cube, then Θ is an equivalence relation and let E_e be a Θ-class. Let $\{G_e^-, G_e^+\}$ be the complementary halfspaces of G defined by setting $G_e^- := G(W(u,v))$ and $G_e^+ := G(W(v,u))$ for an arbitrary edge $uv \in E_e$. An *elementary restriction* consists of taking one of the halfspaces G_e^- and G_e^+. A *restriction* is a convex subgraph of G induced by the intersection of a set of halfspaces of G. Since any convex subgraph of a partial cube G is the intersection of halfspaces [1, 2, 13], the restrictions of G coincide with the convex subgraphs of G. Denote by $\pi_e(G)$ an *elementary contraction*, i.e., the graph obtained from G by contracting the edges in E_e. For a vertex v of G, let $\pi_e(v)$ be the image of v under the contraction. We apply π_e to subsets $S \subseteq V$, by setting $\pi_e(S) := \{\pi_e(v) : v \in S\}$. By [14, Theorem 3], the class of partial cubes is closed under contractions. Since contractions commute, for a set A of Θ-classes, we denote by $\pi_A(G)$ the isometric subgraph of $Q_{m-|A|}$ obtained from G by contracting the equivalence classes of edges from A. Contractions and restrictions also commute in partial cubes. A *pc-minor* of G is a partial cube obtained from G by restrictions and contractions. Since tope graphs of COMs and OMs are partial cubes, we can describe deletions and contractions on sign-vectors in terms of partial cubes.

Lemma 2. *Let $\mathcal{M} = (U, \mathcal{L})$ be a COM and $A \subseteq U$. Then $\pi_A(G(\mathcal{M}))$ is the tope graph of $\mathcal{M} \setminus A$. In particular, if $X \in \mathcal{L}$, then the tope graph of $\mathcal{M}(X) = \mathcal{M} \setminus \underline{X}$ is isomorphic to $[X]$.*

Recall also the following lemma from [4] and [46]:

Lemma 3. *For each covector X of a COM $\mathcal{M}$, the subgraph $[X]$ of the tope graph $G(\mathcal{M})$ is gated.*

2.5. **VC-dimension.** Let $\mathcal{S}$ be a family of subsets of an m-element set U. A subset X of U is *shattered* by $\mathcal{S}$ if for all $Y \subseteq X$ there exists $S \in \mathcal{S}$ such that $S \cap X = Y$. The *Vapnik-Chervonenkis dimension* (VC-dimension) [68] VC-dim($\mathcal{S}$) of $\mathcal{S}$ is the cardinality of the largest subset of U shattered by $\mathcal{S}$. Any set family $\mathcal{S} \subseteq 2^U$ can be viewed as a subset of vertices of the m-dimensional hypercube $Q_m = Q(U)$. Denote by $G(\mathcal{S})$ the *1-inclusion graph* of $\mathcal{S}$, i.e., the subgraph of Q_m induced by the vertices of Q_m corresponding to $\mathcal{S}$. A subgraph G of Q_m has VC-dimension d if G is the 1-inclusion graph of a set family of VC-dimension d. For partial cubes, the VC-dimension can be formulated in terms of pc-minors: a partial cube G has VC-dimension $\leq d$ if and only if G does not have Q_{d+1} as a pc-minor. This is well-defined, since the embeddings of partial cubes are unique up to isomorphism, see e.g. [61, Chapter 5].

The *VC-dimension* VC-dim($\mathcal{M}$) of a COM $\mathcal{M} = (U, \mathcal{L})$ is the VC-dimension of its tope graph $G(\mathcal{M})$. The *VC-dimension* VC-dim(X) of a covector $X \in \mathcal{L}$ of $\mathcal{M}$ is the VC-dimension of the OM $\mathcal{M}(X)$, i.e., it is the VC-dimension of the graph $[X]$. The VC-dimension of OMs, COMs, and their covectors can be expressed in the following way:

Lemma 4. *[19, Lemma 13] For a COM $\mathcal{M}$,* $\text{VC-dim}(\mathcal{M}) = \max\{\text{VC-dim}(\mathcal{M}(X)) : X \in \mathcal{L}\}$. *If $\mathcal{M}$ is an OM and X a cocircuit of $\mathcal{M}$, then* $\text{VC-dim}(X) + 1 = \text{VC-dim}(\mathcal{M}) = \text{rank}(\mathcal{M})$.

That $\text{VC-dim}(X) = \text{VC-dim}(\mathcal{M}) - 1$ for cocircuits X of an OM $\mathcal{M}$ follows from the fact that the cocircuits are atoms of the big face lattice $\mathcal{F}_{\text{big}}(\mathcal{M})$ and this lattice is graded.

3. Auxiliary results

We establish and recall several auxiliary results about OMs and COMs. We also develop a correspondence between realizable samples and convex subgraphs of partial cubes.

3.1. **More about shattering in OMs and COMs.** We continue with new results about shattering in OMs and COMs. Let G be a partial cube, H a convex subgraph, and E_e a Θ-class of G. We say that E_e *crosses* H if H contains an edge of E_e. If E_e does not cross H and there exists an edge uv of E_e with $u \in H$ and $v \notin H$, then E_e and H *osculate*. Otherwise, E_e is *disjoint* from H. Denote by $\operatorname{osc}(H)$ the set of all e such that E_e osculates with H and by $\operatorname{cross}(H)$ the set of all e such that E_e crosses H.

Lemma 5. *Let G be a partial cube, H a convex subgraph, and $e \notin \operatorname{osc}(H)$. Then $\pi_e(H)$ is convex in $\pi_e(G)$ and $\operatorname{osc}(\pi_e(H)) = \operatorname{osc}(H)$, where $\operatorname{osc}(H)$ and $\operatorname{osc}(\pi_e(H))$ are considered in G and $\pi_e(G)$, respectively.*

Proof. Let $H' = \pi_e(H)$. First, since $e \notin \operatorname{osc}(H)$, the fact that H' is a convex subgraph of $\pi_e(G)$ comes from [17, Lemma 5]. Then, the inclusion $\operatorname{osc}(H) \subseteq \operatorname{osc}(H')$ is obvious. If there exists $e' \in \operatorname{osc}(H') \setminus \operatorname{osc}(H)$, then there exists an edge $\pi_e(u)\pi_e(v)$ in $\pi_e(E_{e'})$ with $\pi_e(u) \in V(H')$ and $\pi_e(v) \notin V(H')$. Then $\pi_e(u)\pi_e(v)$ comes from an edge uv of G belonging to $E_{e'}$. Since $e' \notin \operatorname{osc}(H)$, the vertices u and v do not belong to H. This implies that there exists an edge uw of E_e with $w \in V(H)$. Then E_e and H osculate, a contradiction. □

Lemma 6. *Let G be a partial cube and H a gated subgraph of G. If $D \subseteq \operatorname{cross}(H)$ is shattered by G, then D is shattered by H.*

Proof. Pick any Θ-class E_e with $e \in D$ and let v be any vertex of G. If v belongs to the halfspace G_e^- of G, then the gate v' of v in H also belongs to G_e^-. Indeed, since E_e crosses H, there exists a vertex $w \in G_e^- \cap H$. Then $v' \in I(v,w) \subset G_e^-$ by convexity of G_e^- and since v' is the gate of v in H. Analogously, if $v \in G_e^+$, then $v' \in G_e^+$.

Since G shatters D, for any sign vector $X \in \{-1,+1\}^D$, there exists a vertex v_X of G, whose restriction to D coincides with X. This means that for any $e \in D$, the vertex v_X belongs to the halfspace $G_e^{X_e}$. Since the gate v'_X of v_X in H also belongs to $G_e^{X_e}$, the restriction of v'_X to D also coincides with X. This implies that H also shatters D. □

The next lemma shows that the sets shattered by an OM $\mathcal{M}$ are exactly the *independent sets* of the underlying matroid $\underline{\mathcal{M}}$, i.e., the sets not containing supports of circuits of $\mathcal{M}$.

Lemma 7. [1] *Let $\mathcal{M} = (U, \mathcal{L})$ be an* OM *and D be a subset of U. Then D is shattered by $\mathcal{M}$ if and only if D is independent in the underlying matroid $\underline{\mathcal{M}}$.*

Proof. First suppose that D is not shattered by $\mathcal{M}$. We assert that there is a circuit of $\mathcal{M}$ with support included in D, showing that D is not an independent set of $\underline{\mathcal{M}}$. Let $|D| = d+1$. We proceed by induction on $|U| + |D|$. Let $\mathcal{M}' := \mathcal{M} \setminus (U \setminus D)$. Since $G(\mathcal{M})$ does not shatter D and $G(\mathcal{M}')$ is a pc-minor of $G(\mathcal{M})$, $G(\mathcal{M}')$ also does not shatter D. Therefore, if D is a proper subset of U, then by induction hypothesis, there exists a circuit Y' of $\mathcal{M}'$ with $\underline{Y'} \subseteq D$. Consider the sign vector $Y \in \{\pm 1, 0\}^U$ defined by setting $Y_e = Y'_e$ if $e \in D$ and $Y_e = 0$ if $e \in U \setminus D$. By Lemma 1, for any cocircuit X of $\mathcal{M}$, $X' = X \setminus (U \setminus D)$ is a cocircuit of $\mathcal{M}'$. Since $X' \perp Y'$, we conclude that $X \perp Y$. By Theorem 1, Y belongs to $\mathcal{V}$, thus Y contains a circuit of $\mathcal{M}$. Since the support of Y is contained in D, the support of this circuit is also contained in D, and we are done. Therefore, we can suppose that $U = D$.

If D contains a proper subset D', which is not shattered by $\mathcal{M}$, then we can apply the induction hypothesis and find a circuit Y with $\underline{Y} \subseteq D' \subset D$, and we are done. Thus we can suppose that all proper subsets of D are shattered by $\mathcal{M}$. Since $U = D$, this implies that VC-dim$(\mathcal{M}) = |D| - 1 = d$. By Lemma 4, VC-dim$(X) = d-1$ for any cocircuit X of $\mathcal{M}$.

If $\mathcal{M}$ contains a cocircuit X such that $\mathcal{M}(X)$ does not shatter the set $X^0 \cap D$, then by induction assumption applied to $\mathcal{M}(X)$ we can find a circuit Y' of $\mathcal{M}(X)$ with $\underline{Y'} \subseteq X^0 \cap D$. Then extending

[1]We would like to acknowledge Emeric Gioan for finding a gap in the proof of this lemma.

Y' to Y as in the previous case, we obtain a circuit Y of $\mathcal{M}$ whose support is included in D and we are done. Therefore, we can suppose that for any cocircuit X of $\mathcal{M}$, $\mathcal{M}(X)$ shatters the set $X^0 \cap D$. Since VC-dim$(X) = d-1$ and $U = D$, this implies that $|X^0 \cap D| = d-1$. Hence, the support $\underline{X}$ of each cocircuit X consists of two elements.

Since D is not shattered by $\mathcal{M}$, there exists a sign-vector $Y' \in \{-1,+1\}^D$ such that for any tope T of $\mathcal{M}$, the restriction of T to D is different from Y'. By symmetry, $-Y'$ also is not shattered by $\mathcal{M}$. Consider the sign vector $Y \in \{\pm 1, 0\}^U$ defined by setting $Y_e = Y'_e$ if $e \in D$ and $Y_e = 0$ if $e \in U \setminus D$. Then $Y, -Y \notin \mathcal{L}$. We assert that Y is a vector of $\mathcal{M}$. By Theorem 1, we have to show that $Y \perp X$ for any cocircuit X of $\mathcal{M}$. We assert that $X_f Y_f = -X_{f'} Y_{f'}$, where $\underline{X} = \{f, f'\}$. Indeed, since Y and $-Y$ do not belong to $\uparrow X$, $\text{Sep}(Y, -Y) = D$, and $X^0 = D \setminus \{f, f'\}$, we must have $X_f \neq Y_f, X_{f'} \neq -Y_{f'}$ or $X_{f'} \neq Y_{f'}, X_f \neq -Y_f$. In the first case we have $X_f Y_f = -1, X_{f'} Y_{f'} = +1$ and in the second case we have $X_f Y_f = +1, X_{f'} Y_{f'} = -1$. This proves that $Y \in \mathcal{V}$, thus the support of Y contains a circuit of $\mathcal{M}$. Since $Y_e = 0$ if $e \in U \setminus D$, we deduce that $\underline{Y} \subseteq D$. Since $\underline{Y}$ contains a circuit of $\mathcal{M}$, D is not an independent set of $\underline{\mathcal{M}}$.

Conversely, let D be a set of size d shattered by $\mathcal{M}$ and suppose by way of contradiction that D contains a circuit of $\underline{\mathcal{M}}$. Since any subset of D is also shattered by $\mathcal{M}$, we can suppose without loss of generality that for any $e \in D$, $D \setminus \{e\}$ is an independent set of $\underline{\mathcal{M}}$, i.e. that D is a circuit of $\underline{\mathcal{M}}$. By passing from $\mathcal{M}$ to $\mathcal{M}' = \mathcal{M} \setminus (U \setminus D)$, we can also suppose that $U = D$, i.e., that VC-dim$(\mathcal{M}) = d$. Since $\mathcal{M}$ shatters the set $D = U$, any sign vector from $\{\pm 1\}^D$ is a tope of $\mathcal{M}$. Consider the two circuit signatures Y and $-Y$ of the circuit D. They are obviously topes of $\mathcal{M}$. Consider also a cocircuit X of $\mathcal{M}$ such that $-Y \in \uparrow X$ and $Y \notin \uparrow X$ (such a X exists because $\mathcal{M}$ is a simple OM). By Lemma 4 VC-dim$(X) = d-1$, thus D contains an element f such that X shatters $D \setminus \{f\}$. This implies that $D \setminus \{f\} \subseteq X^0$. Hence, Y and X are not orthogonal, contrary to Theorem 1. This shows that each set D shattered by $\mathcal{M}$ is an independent set of $\underline{\mathcal{M}}$. $\square$

An *antipode* of a vertex v in a partial cube G is a (necessarily unique) vertex $-v$ such that $G = I(v, -v)$. A partial cube G is *antipodal* if all its vertices have antipodes. By (Sym), a tope graph of a COM is the tope graph of an OM if and only if it is antipodal, see [46].

The next lemma can be seen as dual analogue of Lemma 7. It shows that the VC-dimension of OMs is defined locally at each tope T, by shattering subsets of $\text{osc}(T)$.

Lemma 8. *Let $\mathcal{M} = (U, \mathcal{L})$ be an* OM *of rank d with tope graph $G(\mathcal{M})$. For any tope T of $\mathcal{M}$, $\text{osc}(T)$ contains a subset D of size d shattered by $\mathcal{M}$.*

Proof. We proceed by induction on the size of U. If $\text{osc}(T) = U$, then we are obviously done. Thus suppose that there exists $e \notin \text{osc}(T)$. Consider the tope graph $G' = \pi_e(G)$ of the oriented matroid $\mathcal{M}' = \mathcal{M} \setminus e$. Let $T' = \pi_e(T)$. Then $\text{osc}(T') = \text{osc}(T)$ by Lemma 5. If $\text{rank}(\mathcal{M}') = d$, by induction hypothesis the set $\text{osc}(T')$ contains a subset D of size d shattered by G'. Since G' is a pc-minor of G, $D \subset \text{osc}(T)$ is also shattered by G and we are done.

Thus, let $\text{rank}(\mathcal{M}') < \text{rank}(\mathcal{M})$. If the Θ-class E_e of G crosses the faces $\uparrow X$ of all cocircuits $X \in \mathcal{L}$, then $\mathcal{L}$ is not simple. Therefore, there exists a cocircuit $X \in \mathcal{L}$ whose face $\uparrow X$ is not crossed by E_e. However, since when we contract E_e the rank decreases by 1, the resulting OM $\mathcal{M}'$ coincides with $\uparrow X$. Indeed, after contraction the rank of $\uparrow X$ remains the same. Hence, if X would remain a cocircuit, then the global rank would not decrease. Hence, G' is the tope graph of $\mathcal{M}(X)$. Since G is an antipodal partial cube and $G_e^+ = \uparrow X$, we have $G_e^- \cong G_e^+$. This shows that $G \cong G_e^+ \square K_2 \cong G' \square K_2$. This implies that E_e osculate with $\{T\}$ in G, contrary to the assumption $e \notin \text{osc}(T)$. $\square$

Next we give a shattering property of COMs. It is in its proof where the axiom (SE) of COMs is crucial. The *distance* $d(A, B)$ between sets A, B of vertices of G is $\min\{d(a, b) : a \in A, b \in B\}$. The set $\text{pr}_B(A) = \{a \in A : d(a, B) = d(A, B)\}$ is the *metric projection* of B on A. For two covectors $X, Y \in \mathcal{L}$ of a COM $\mathcal{M}$, we denote by $\text{pr}_{[X]}([Y])$ the metric projection of $[X]$ on $[Y]$ in $G(\mathcal{M})$. Since

$[X]$ and $[Y]$ are gated by Lemma 3, $\mathrm{pr}_{[X]}([Y])$ consists of the gates of vertices of $[X]$ in $[Y]$. Two faces $\uparrow X$ and $\uparrow Y$ of $\mathcal{M}$ are *parallel* if $\mathrm{pr}_{[X]}([Y]) = [Y]$ and $\mathrm{pr}_{[Y]}([X]) = [X]$. A *gallery* between two parallel faces $\uparrow X$ and $\uparrow Y$ of $\mathcal{M}$ is a sequence of faces $(\uparrow X = \uparrow X_0, \uparrow X_1, \ldots, \uparrow X_{k-1}, \uparrow X_k = \uparrow Y)$ such that any two faces of this sequence are parallel and any two consecutive faces $\uparrow X_{i-1}, \uparrow X_i$ are facets of a common face of $\mathcal{L}$. A *geodesic gallery* between $\uparrow X$ and $\uparrow Y$ is a gallery of length $|\mathrm{Sep}(X,Y)|$. Two parallel faces $\uparrow X, \uparrow Y$ are *adjacent* if $|\mathrm{Sep}(X,Y)| = 1$, i.e., $\uparrow X$ and $\uparrow Y$ are opposite facets of a face of $\mathcal{L}$. See Figure 3 and recall the following result:

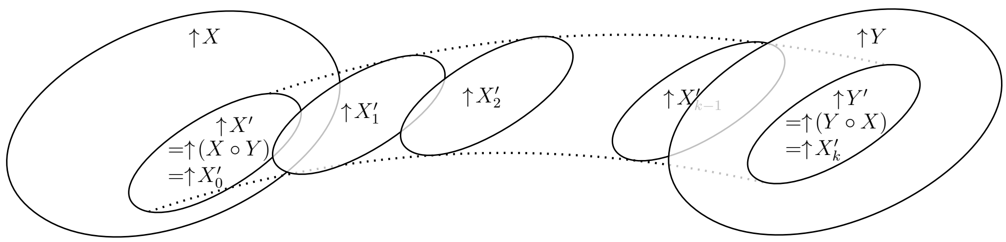

FIGURE 3. Illustration of Lemmas 9 and 10.

Lemma 9. *[19, Proposition 8] Let $\mathcal{M} = (U, \mathcal{L})$ be a COM and $X, Y \in \mathcal{L}$. Then:*

(i) *$d([X],[Y]) = |S(X,Y)|$ and the gates of $[Y]$ in $[X]$ are the vertices of $[X \circ Y] \subseteq [X]$;*

(ii) *$\uparrow X$ and $\uparrow Y$ are parallel if and only if $\underline{X} = \underline{Y}$. If $\uparrow X$ and $\uparrow Y$ are parallel, then they are connected by a geodesic gallery;*

(iii) *$\mathrm{pr}_{[Y]}([X]) = [X \circ Y]$, $\mathrm{pr}_{[X]}([Y]) = [Y \circ X]$, and $\uparrow(X \circ Y)$ and $\uparrow(Y \circ X)$ are parallel.*

A covector $X \in \mathcal{L}$ of a COM $\mathcal{M} = (U, \mathcal{L})$ *maximally shatters* a set $D \subseteq U$ if $[X]$ shatters D but $[X]$ does not shatter any superset of D. We also say that $X \in \mathcal{L}$ *minimally shatters* a set D if $[X]$ shatters D but D is not shattered by $[X']$ for any covector $X' > X$.

Lemma 10. *Let $\mathcal{M} = (U, \mathcal{L})$ be a COM and $X, Y \in \mathcal{L}$. Then:*

(i) *if $[X]$ and $[Y]$ shatter D, then the projections $[X \circ Y]$ and $[Y \circ X]$ also shatter D;*

(ii) *if $[X]$ maximally shatters D and $[Y]$ shatters D, then $[X \circ Y] = [X]$ and $\uparrow X$ is not a facet of $\mathcal{M}$;*

(iii) *if both $[X]$ and $[Y]$ shatter D, then there exist covectors $X' \geq X, Y' \geq Y$ such that $[X']$ and $[Y']$ both maximally shatter D, and $\uparrow X'$ and $\uparrow Y'$ are parallel.*

Proof. **Property (i):** Since $[X]$ and $[Y]$ shatter D, for any sign vector $Z \in \{\pm 1\}^D$ we can find two topes $T' \in [X]$ and $T'' \in [Y]$, such that $T'_{|D} = Z = T''_{|D}$. Since $X < T'$ and $Y < T''$, from $T'_{|D} = Z = T''_{|D}$ we conclude that $(X \circ Y)_{|D} < Z$ and in $[X \circ Y]$ we can find a tope T whose restriction to D coincides with Z. This proves that $[X \circ Y]$ shatters D, establishing (i).

Property (ii): If $[X]$ max-shatters D, then VC-dim$(X) = |D| =: d$. By property (i), $[X \circ Y]$ also shatters D. If $\uparrow(X \circ Y)$ is a proper face of $\uparrow X$, then we obtain a contradiction with Lemma 4. Thus $\uparrow(X \circ Y) = \uparrow X$, showing that $X = X \circ Y$. This establishes the first assertion. By Lemma 9, the faces $\uparrow X$ and $\uparrow(Y \circ X)$ are parallel and therefore are connected by a geodesic gallery $(\uparrow X = \uparrow X_0, \uparrow X_1, \ldots, \uparrow X_k = \uparrow(Y \circ X))$. Then $\uparrow X$ and $\uparrow X_1$ are facets of a common face of $\mathcal{L}$, thus $\uparrow X$ is not a facet of $\mathcal{M}$. This proves (ii).

Property (iii): Let $d = |D|$. We can suppose that both X and Y minimally shatter the set D. Indeed, if D is shattered by a proper face $\uparrow X'$ of $\uparrow X$, then we can replace the pair X, Y by the pair X', Y so that $[X']$ and $[Y]$ still shatter D. Thus D is not shattered by any proper faces of $\uparrow X$ and $\uparrow Y$. Since by (i), D is shattered by $[X \circ Y]$ and $[Y \circ X]$, we conclude that $X = X \circ Y$ and $Y = Y \circ X$ and thus the faces $\uparrow X$ and $\uparrow Y$ are parallel.

It remains to show that $[X]$ and $[Y]$ max-shatter D. Suppose by way of contradiction that $[X]$ shatters a larger set $D' := D \cup \{e\}$. Consider the OM $\mathcal{M}' = \uparrow X \setminus (U \setminus D')$. Since $[X]$ shatters D', $\mathcal{M}'$ also shatters D'. Moreover, $\mathcal{M}'$ max-shatters D', i.e., VC-dim$(\mathcal{M}') = d + 1$. Since $\mathcal{M}'$ is a simple OM, $\mathcal{M}'$ contains two adjacent topes T'_1, T'_2 with $\mathrm{Sep}(T'_1, T'_2) = \{e\}$ and we can find a cocircuit X'' of $\mathcal{M}'$ such that $T'_1 \in [X'']$ and $T'_2 \notin [X'']$. By Lemma 4 applied to $\mathcal{M}'$, we conclude that X'' has VC-dimension d. Hence, X'' must shatter the set D. By Lemma 1, there is a cocircuit X' of $\uparrow X$ such that $X'' = X' \setminus (U \setminus D')$. Since X'' shatters D, X' also shatters D. Since $X < X'$, this contradicts our assumption that X minimally shatters D. This establishes (iii). □

3.2. **Realizable and full samples as convex subgraphs.** Let $\mathcal{L} \subset \{-1, 0, +1\}^U$ be a system of sign vectors whose topes $\mathcal{T}$ induce an isometric subgraph G of $\mathrm{Q}(U)$. Recall that $\downarrow\mathcal{L} = \bigcup_{X \in \mathcal{L}} \{S \in \{-1, 0, +1\}^U : S \leq X\}$ is the *set of realizable samples* for $\mathcal{L}$ (this is called *polar complex* in neural codes [45]). Since for any $X \in \mathcal{L}$ there exists $T \in \mathcal{T}$ such that $X \leq T$, we have $\downarrow\mathcal{L} = \downarrow\mathcal{T}$, see Figure 4. For a realizable sample $S \in \downarrow\mathcal{L}$, let $\uparrow S = \{X \in \mathcal{L} : S \leq X\}$ and let $[S]$ be the subgraph of G induced by all topes $T \in \mathcal{L}$ from $\uparrow S$. For OMs, the set $\uparrow S$ is called *supertope* in [44]. For COMs, $\uparrow S$ is called the *fiber* of S and it is known that they are COMs [4]. Since for any $S \in \downarrow\mathcal{L}$ there exists $T \in \mathcal{T}$ such that $S \leq T$, $[S] \neq \emptyset$. Moreover, $[S]$ is the intersection of halfspaces of G of the form G^+_e if $S_e = +1$ and G^-_e if $S_e = -1$. Hence, $[S]$ is a nonempty convex subgraph of G for all $S \in \downarrow\mathcal{L}$. If $\mathcal{M} = (U, \mathcal{L})$ is a COM and $S \in \mathcal{L}$, then $\uparrow S$ is the face of S and $[S]$ is gated by Lemma 3.

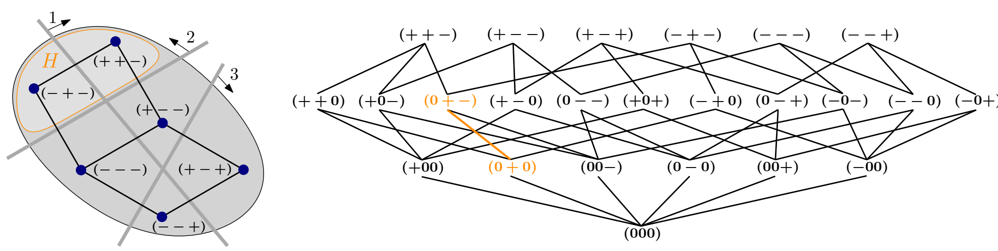

FIGURE 4. Left: the tope graph G of the restriction $\mathcal{M}$ of the COM from Figure 1 to $\{1, 2, 3\}$ and a convex subgraph H of G. Right: the realizable samples of $\mathcal{M}$ and the interval $I(H)$ (in orange).

Any convex subgraph H of a partial cube G is the intersection of all halfspaces of G containing H. However, H can be represented in different ways as the intersection of halfspaces. Indeed, any representation of H as an intersection of halfspaces of G yields a realizable sample S, where $S_e = \pm 1$ if $G^{\pm}_e$ participates in the representation and $S_e = 0$ otherwise. Notice that the Θ-classes osculating with H have to be part of every representation of H and the Θ-classes crossing H take part in no representation of H. This leads to two canonical representations of H, one using only the halfspaces whose Θ-class osculates with H and one using all halfspaces containing H:

$$(S_\perp)_e = \begin{cases} -1 & \text{if } e \in \mathrm{osc}(H) \text{ and } H \subseteq G^-_e, \\ +1 & \text{if } e \in \mathrm{osc}(H) \text{ and } H \subseteq G^+_e, \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad (S^\top)_e = \begin{cases} -1 & \text{if } H \subseteq G^-_e, \\ +1 & \text{if } H \subseteq G^+_e, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $(S^\top)^0 = \mathrm{cross}(H)$ and $(S_\perp)^0 = U \setminus \mathrm{osc}(H)$, i.e., $(S_\perp)^0$ consists of all e such that E_e crosses or is disjoint from H. If S is a sample arising from the representation of H as the intersection of halfspaces, then $S_\perp \leq S \leq S^\top$. Moreover, any sample S from the order interval $I(H) := [S_\perp, S^\top]$ arises from a representation of H, i.e., $[S] = [S_\perp] = [S^\top] = H$. Thus, for any convex subgraph H of G the set of all $S \in \downarrow\mathcal{L}$ such that $[S] = H$ is an interval $I(H) = [S_\perp, S^\top]$ of $(\downarrow\mathcal{L}, \leq)$. Note that the intervals $I(H)$ partition $\downarrow\mathcal{L}$. Moreover:

Lemma 11. *If $S, S' \in \downarrow\mathcal{L}$ and $S \leq S'$, then $[S'] \subseteq [S]$.*

Lemma 12. *If* $X \in \mathcal{L}, S \in \downarrow\mathcal{L}$ *such that* $\mathrm{Sep}(X, S) = \varnothing$ *and* $\widehat{S} = X \circ S$, *then* $[\widehat{S}] = [X] \cap [S]$.

Proof. Since $\widehat{S} = X \circ S$, we have $X \leq \widehat{S}$ and by Lemma 11 we have $[\widehat{S}] \subseteq [X]$. Now we prove that $[\widehat{S}] \subseteq [S]$. Indeed, otherwise there exists a tope T of $\mathcal{L}$ such that $T \in [\widehat{S}] \setminus [S]$. This implies that $\widehat{S} < T$ and there exists an element $e \in U$ such that $T_e \neq S_e \neq 0$. Since $\widehat{S} = X \circ S$, this implies that $X_e = T_e$, which is impossible because $\mathrm{Sep}(X, S) = \varnothing$. This proves that $[\widehat{S}] \subseteq [X] \cap [S]$.

To prove the converse inclusion $[X] \cap [S] \subseteq [\widehat{S}]$ pick any tope T of $\mathcal{L}$ belonging to $[X] \cap [S]$. Then $X < T$ and $S < T$. Suppose by way of contradiction that $T \notin [\widehat{S}]$, i.e., $\widehat{S} \not< T$. Then there exists $e \in U$ such that $\widehat{S}_e \neq 0$ and $\widehat{S}_e \neq T_e$, say $\widehat{S}_e = -1$ and $T_e = +1$. Since $\widehat{S} = X \circ S$, the equality $\widehat{S}_e = -1$ implies that either $X_e = -1$ or that $X_e = 0$ and $S_e = -1$. Since $T_e = +1$, in the first case we get a contradiction with $X < T$ and in the second case we get a contradiction with $S < T$. Hence, $[X] \cap [S] \subseteq [\widehat{S}]$ and we are done. □

We say that a sample $S \in\downarrow \mathcal{L}$ is *full* if the pc-minor $G' = \pi_{S^0}(G)$ has VC-dimension $d = \text{VC-dim}(G)$. Let $\downarrow \mathcal{L}_f$ denote the set of all full samples of $\mathcal{L}$. Note that all topes of $\mathcal{L}$ are full samples since their zero set is empty. A convex subgraph H of G is *full* if $S_\perp(H)$ is full. The image of H in G' is a single vertex v_H and its degree is $|\mathrm{osc}(H)|$. If $D \subset \mathrm{osc}(v_H) = \mathrm{osc}(H)$ of size d is shattered by G', since G' is a pc-minor of G, D is also shattered by G. Hence, a convex set H of G is full if and only if G shatters a subset D of $\mathrm{osc}(H)$ of size $d = \text{VC-dim}(G)$. If H is a full convex subgraph of a COM, not all samples in $I(H)$ have to be full. We show next, that the above problem does not arise in OMs.

Lemma 13. *Let* $\mathcal{M} = (U, \mathcal{L})$ *be an OM of rank* d *and let* $G = G(\mathcal{M})$ *be its tope graph. A sample* $S \in\downarrow\mathcal{L}$ *is full if and only if the convex subgraph* $[S]$ *is full.*

Proof. First notice that since in OMs the rank and the VC-dimension are equal, a sample S is full if and only if $\mathrm{rank}(\mathcal{M} \setminus S^0) = d$.

First suppose that the convex subgraph H is full. Then the sample $S_\perp := S_\perp(H)$ is full. Since $\underline{S_\perp} = \mathrm{osc}(H) \subseteq \underline{S}$ for any $S \in I(H) = [S_\perp, S^\top]$, we get $S^0 \subset (S_\perp)^0$, thus $\mathrm{rank}(\mathcal{M} \setminus S^0) \geq \mathrm{rank}(\mathcal{M} \setminus (S_\perp)^0) = d = \mathrm{rank}(\mathcal{M})$. Hence, $\mathrm{rank}(\mathcal{M} \setminus S^0) = d$, i.e., S is a full sample.

Conversely, let S be a full sample and we assert that $H = [S]$ is a full convex subgraph. Let $\mathcal{M}' = \mathcal{M} \setminus \mathrm{cross}(H)$ and let $G' = \pi_{\mathrm{cross}(H)}(G)$ be its tope graph. Since $\mathrm{cross}(H) \subseteq S^0$ and S is full, $\mathrm{rank}(\mathcal{M}') = d$ and hence $\text{VC-dim}(G') = d$. The image of H in G' is a single vertex v_H. By Lemma 5, $\mathrm{osc}(v_H) = \mathrm{osc}(H)$. By Lemma 8, $\mathrm{osc}(v_H)$ contains a subset of size d shattered by $\mathcal{M}'$, whence H is full. □

4. The main result

The goal of this section is to prove the following theorem:

Theorem 3. *The set* $\mathcal{T}$ *of topes of a complex of oriented matroids* $\mathcal{M} = (U, \mathcal{L})$ *of VC-dimension* d *admits a proper labeled sample compression scheme of size* d.

4.1. **The main idea.** Our labelled sample compression scheme takes any realizable sample S of a COM $\mathcal{M}$ and removes the zero-set of S. Consequently, S becomes the tope $S' = S \setminus S^0$ of the COM $\mathcal{M} \setminus S^0 =: \mathcal{M}'$. Then we consider a face $\uparrow X'$ of $\mathcal{M}'$ defined by a minimal covector X' of $\mathcal{M}'$ such that $S' \geq X'$. This face is the OM $\mathcal{M}'(X') = \uparrow X' \setminus \underline{X'}$. The compressor $\alpha(S)$ is then defined by applying to S' and $\mathcal{M}'(X')$ the *distinguishing lemma*, which allows to distinguish full samples of an OM of rank d by considering their restriction to subsets of size d. It constructs a function that assigns such a subset to each full sample and is used by both compressor and reconstructor. The *localization lemma* is used by the reconstructor and designates the set of all potential covectors whose faces may contain topes T compatible with the initial sample S. These two lemmas are proved in next two subsections. Compressor and reconstructor are given in the last subsection and are illustrated by Example 1. The compressor generalizes the compressor for ample classes of Moran and Warmuth [56]. However, the reconstructor is much more involved than that for ample classes.

4.2. **The distinguishing lemma.** In this subsection, $\mathcal{M} = (U, \mathcal{L})$ is an OM of rank d. The *distinguishing lemma* allows to distinguish full samples of $\mathcal{M}$ by considering their restriction to subsets of size d. It is constructing a function $f_{\mathcal{M}}$ that assigns such a subset to each full sample and is used by both compressor and reconstructor.

Lemma 14. *Let $\mathcal{M} = (U, \mathcal{L})$ be an* OM *of VC-dimension d. Then there exists a function $f_{\mathcal{M}} :\downarrow \mathcal{L}_f \to \binom{U}{d}$ such that for all $S, S' \in\downarrow\mathcal{L}_f$:*

(i) *if $e \in f_{\mathcal{M}}(S)$, then $e \in \mathrm{osc}([S])$,*
(ii) *$f_{\mathcal{M}}(S)$ is shattered by $\mathcal{M}$,*
(iii) *if $e \notin \mathrm{osc}([S])$, then $f_{\mathcal{M}}(S) = f_{\mathcal{M}\setminus e}(S \setminus e)$,*
(iv) *if $S_{|f_{\mathcal{M}}(S)} = S'_{|f_{\mathcal{M}}(S')}$, then $[S] = [S']$.*

Proof. Let $G := G(\mathcal{M})$ be the tope graph of $\mathcal{M}$. We proceed by induction on d. If $d = 1$, then $U = \{e\}$ and G is an edge between the topes $T_1 = (-1)$ and $T_2 = (+1)$, which are the only full samples of $\mathcal{M}$ Defining $f_{\mathcal{M}}(T_1) = f_{\mathcal{M}}(T_2) = \{e\}$, we obtain a function satisfying the conditions (i)-(iv). Before treating the general case $d \geq 2$, we establish the following claim:

Claim 1. If S is a full sample and $e \in \mathrm{osc}([S])$, then there exists a cocircuit X of $\mathcal{M}$ such that $e \in \underline{X}$ and $X \leq S$. Moreover, $S \setminus \underline{X}$ is a full sample of $\mathcal{M}(X)$.

Proof. Since S is a full sample, $\mathcal{M}' = \mathcal{M} \setminus S^0$ has VC-dimension d and thus rank d. Moreover, $[S \setminus S^0]$ is a tope T of $\mathcal{M}'$. By Lemma 5, $e \in \mathrm{osc}([S]) = \mathrm{osc}([T])$ and T is incident to an edge in E_e, i.e., there is a tope T' of $\mathcal{M}'$ such that $\mathrm{Sep}(T, T') = \{e\}$. Let X' be a cocircuit of $\mathcal{M}'$ such that its face $\uparrow X'$ contains T but not T'. This cocircuit X' exists, otherwise all cocircuits Y' of $\mathcal{M}'$ would have $Y'_e = 0$, but $\mathcal{M}'$ is simple because $\mathcal{M}$ is. Now, since $\mathcal{M}'$ has VC-dimension d, $\mathcal{M}'(X') \cong\uparrow X'$ has VC-dim $d-1$ by Lemma 4. Thus, there exists a covector X of $\mathcal{M}$ such that $X' = X \setminus S^0$ and VC-dim$(X) = d - 1$. If X is not a cocircuit, then $\uparrow X$ is a proper face of $\uparrow Y$ for a cocircuit Y of $\mathcal{M}$. Since the VC-dimension of any proper face is strictly smaller than the VC-dimension of the face itself and since $\mathcal{M}$ has rank d, we obtain a contradiction. Thus X is a cocircuit of $\mathcal{M}$. In particular, $e \in \underline{X}$ and $X \leq S$. It remains to show that $S \setminus \underline{X}$ is a full sample of $\mathcal{M}(X)$, i.e., that the VC-dimension of $(\uparrow X \setminus \underline{X}) \setminus S^0$ is $d-1$. To see this note that by Lemma 1, $X \setminus S^0 \in \mathcal{C}^*(\mathcal{M} \setminus S^0)$ and since S is full the VC-dimension of $\mathcal{M} \setminus S^0$ is d. Hence, the VC-dimension of $\mathcal{M}(X \setminus S^0)$ is $d-1$ by Lemma 4. But since all sign-vectors in $\uparrow X$ agree on $\underline{X}$, we have $\mathcal{M}(X \setminus S^0) \cong (\uparrow X \setminus \underline{X}) \setminus S^0$. Hence, VC-dim$((\uparrow X \setminus \underline{X}) \setminus S^0) =$ VC-dim$(\mathcal{M}(X \setminus S^0)) =$ VC-dim$(\mathcal{M}) - 1 = d - 1$. □

Let now $d \geq 2$. Fix a linear order on $U = \{1, \ldots, m\}$. For a full sample S, define $f_{\mathcal{M}}(S)$ recursively by setting $f_{\mathcal{M}}(S) = \{e_S, f_{\mathcal{M}(X)}(S \setminus \underline{X})\}$, where e_S is the smallest element of U such that E_{e_S} osculates with $[S]$ and X is any cocircuit of $\mathcal{M}$ such that $e_S \in \underline{X}$ and $X \leq S$. By Claim 1, X exists and by Lemma 11 $[S] \subseteq [X]$ holds. Now we prove that $f_{\mathcal{M}}$ satisfies the conditions (i)-(iv). By Claim 1, $S \setminus \underline{X}$ is a full sample of $\mathcal{M}(X)$ to which we can apply the induction hypothesis.

Condition (i): If $e \in f_{\mathcal{M}}(S)$, then either $e = e_S$ or $e \in f_{\mathcal{M}(X)}(S \setminus \underline{X})$. In the first case, E_e and $[S]$ osculate by the choice of e_S from $\mathrm{osc}([S])$. In the second case, E_e and $[S \setminus \underline{X}]$ osculate in the tope graph of $\mathcal{M}(X)$ by induction hypothesis. Since the tope graph of $\mathcal{M}(X)$ is isomorphic to $[X]$, E_e crosses $[X]$ and thus $e \notin \underline{X}$. Moreover, since $X \leq S$, by Lemma 11, $[S] \subseteq [X]$. Thus $[S \setminus \underline{X}]$ is isomorphic to $[S]$. Hence, E_e and $[S]$ osculate in G.

Condition (ii): Suppose that $f_{\mathcal{M}}(S) = \{e_S, f_{\mathcal{M}(X)}(S \setminus \underline{X})\}$ is not shattered by $\mathcal{M}$. Define $D' = f_{\mathcal{M}(X)}(S \setminus \underline{X})$. By induction hypothesis, D' is shattered by $\mathcal{M}(X)$. Hence, by Lemma 7 there is a circuit Y of $\mathcal{M}$ such that $\underline{Y} \subseteq \{e_S\} \cup D'$ and $e_S \in \underline{Y}$. On the other hand, we have that $D' \subseteq X^0$ and $e_S \in \underline{X}$. Thus, $|\underline{Y} \cap \underline{X}| = 1$, and since X is a cocircuit and Y is a circuit, this contradicts orthogonality of circuits and cocircuits in OMs, see Theorem 1.

Condition (iii): Let $e \notin \mathrm{osc}([S])$. Then clearly $e \notin \mathrm{osc}([S \setminus \underline{X}])$ in $[X \setminus \underline{X}]$. Thus, $e \notin f_{\mathcal{M}}(S)$. Moreover, contracting a class that does not osculate with $[S]$ cannot yield a new class that osculates with $[S]$ by Lemma 5. Thus, by the definition of $f_{\mathcal{M}}$ and induction hypothesis we have $f_{\mathcal{M}}(S) = \{e_S, f_{\mathcal{M}(X)}(S \setminus \underline{X})\} = \{e_S, f_{\uparrow X \setminus (\underline{X} \cup \{e\})}(S \setminus (\underline{X} \cup \{e\}))\} = f_{\mathcal{M} \setminus e}(S \setminus e)$.

Condition (iv): Let S, S' be two full samples such that $S_{|f_{\mathcal{M}}(S)} = S'_{|f_{\mathcal{M}}(S')}$. In particular, $f_{\mathcal{M}}(S) = \{e_S, f_{\mathcal{M}(X)}(S \setminus \underline{X})\} = \{e_{S'}, f_{\mathcal{M}(X')}(S' \setminus \underline{X'})\} = f_{\mathcal{M}}(S')$. By the minimality in the choice of the elements e_S and $e_{S'}$ both are the smallest elements of the respective sets $f_{\mathcal{M}}(S)$ and $f_{\mathcal{M}}(S')$, whence $e_S = e_{S'} =: e$. This means that $f_{\mathcal{M}(X)}(S \setminus \underline{X}) = f_{\mathcal{M}(X')}(S' \setminus \underline{X'}) =: D'$ and for cocircuits X and X' both faces $\uparrow X \cong \mathcal{M}(X)$ and $\uparrow X' \cong \mathcal{M}(X')$ shatter the same set $D' \subseteq U$. By Lemma 10 this implies that $X = X'$ or $X = -X'$. Indeed, let $X \neq X'$. Since X, X' max-shatter D', by Lemma 10(ii) $X = X \circ X'$ and $X' = X' \circ X$. By Lemma 10(iii) there exists a geodesic gallery between $\uparrow X$ and $\uparrow X'$. Since X and X' are cocircuits of $\mathcal{M}$, $\uparrow X$ and $\uparrow X'$ are facets of $\mathcal{M}$. Therefore $\uparrow X$ and $\uparrow X'$ must be consecutive in the gallery and the face containing them as facets must coincide with $\mathcal{M}$. Thus, $X = -X'$.

But $X = -X'$ cannot happen, otherwise, since $e \in \underline{X} \cap \underline{X'}$, we have $e \in \mathrm{Sep}(X, X')$ and since $S \geq X$ and $S' \geq X'$, we get $S_e = -S'_e$, which contradicts the assumption $S_{|f_{\mathcal{M}}(S)} = S'_{|f_{\mathcal{M}}(S')}$. Hence, $X = X'$. Hence $S, S' \geq X$ and by Lemma 11 we get $[S] \subseteq [X]$ and $[S'] \subseteq [X]$. By induction hypothesis, $[S \setminus \underline{X}] = [S' \setminus \underline{X}]$ in $[X]$. This means that $[S] \cap [X] = [S'] \cap [X]$, but since $[S] \subseteq [X]$ and $[S'] \subseteq [X]$, we conclude that $[S] = [S']$. □

4.3. **The localization lemma.** The *localization lemma* designates for any realizable sample of a COM $\mathcal{M}$ the set of all potential covectors whose faces may contain topes of $\mathcal{M}$ which can be used by the reconstructor.

Let $\mathcal{M} = (U, \mathcal{L})$ be a COM of VC-dimension d and let $S \in \downarrow \mathcal{L}$ be a realizable sample. Consider the tope $S' = S \setminus S^0$ of the COM $\mathcal{M}' := \mathcal{M} \setminus S^0$ and let X' be a minimal covector of $\mathcal{M}'$ such that $S' \geq X'$. Note that if $\mathcal{M}'$ is an OM, then $X' = \mathbf{0}$ and $\uparrow X' = \mathcal{M}'$. By Lemma 4, the OM $\mathcal{M}'(X') \cong \uparrow X'$ has VC-dimension $\leq d$. Let

$$\mathcal{H}_{S,X'} := \{X \in \mathcal{L} : X \setminus S^0 = X' \text{ and } \mathcal{M}(X) \text{ has the same VC-dimension as } \mathcal{M}'(X')\}.$$

Let $D \subseteq U \setminus S^0$ be a set of size $d' = \text{VC-dim}(X')$ shattered by the OM $\mathcal{M}'(X')$. Let also

$$\mathcal{H}_D := \{X \in \mathcal{L} : \mathcal{M}(X) \text{ max-shatters } D\}.$$

Lemma 15. *Let $S \in \downarrow \mathcal{L}$, X' be a minimal covector of $\mathcal{M}' = \mathcal{M} \setminus S^0$ such that $S \setminus S^0 = S' \geq X'$, and let $D \subseteq U$ be shattered by $\mathcal{M}'(X') = \uparrow X'$. Then $\varnothing \neq \mathcal{H}_{S,X'} = \mathcal{H}_D$.*

Proof. By Lemma 1 there must be a covector $X \in \mathcal{L}$ such that $X \setminus S^0 = X'$. Moreover, if $\mathcal{M}'(X') = \uparrow X'$ shatters D, then $\mathcal{M}(X)$ also shatters D because the tope graph of $\mathcal{M}'(X')$ is a pc-minor of the tope graph of $\mathcal{M}(X)$. Suppose that $\mathcal{M}(X)$ shatters a superset of D. Then there is a covector $Y \geq X$ of $\mathcal{M}$ such that $\mathcal{M}(Y)$ shatters D. Hence, $Y \setminus S^0 \geq X \setminus S^0 = X'$, but $\mathcal{M}'(Y \setminus S^0)$ and $\mathcal{M}'(X')$ have the same VC-dimension, so by Lemma 4 $Y \setminus S^0 = X'$. Hence $Y \in \mathcal{H}_{S,X'}$. In particular, we have shown that any element of $\mathcal{H}_{S,X'}$ shatters D and $|D|$ is its VC-dimension so it max-shatters D. Hence, $\mathcal{H}_{S,X'} \subseteq \mathcal{H}_D$.

It remains to show $\mathcal{H}_D \subseteq \mathcal{H}_{S,X'}$. Let $Y \in \mathcal{H}_D \setminus \mathcal{H}_{S,X'}$ and set $Y' = Y \setminus S^0$. By assumption we have $X' \neq Y'$ and since $\mathcal{M}(Y)$ max-shatters D and $D \subseteq \underline{S}$, also $\mathcal{M}'(Y')$ max-shatters D. In particular, $D \subseteq X'^0 \cap Y'^0 = (X' \circ Y')^0$. By Lemma 9 the gates of $[Y']$ in $[X']$ are the topes of $\uparrow(X' \circ Y') \subseteq \uparrow X'$. Thus, $[X' \circ Y']$ is a gated subgraph of $[X']$, and $[X' \circ Y']$ is crossed by D, and D is shattered by $[X']$. By Lemma 6, the VC-dimension of $\mathcal{M}'(X' \circ Y')$ is at least $|D|$, which is the VC-dimension of $\mathcal{M}'(X')$. Lemma 4 yields $X' \circ Y' = X'$.

If $\mathrm{Sep}(X', Y') = \varnothing$, then $\uparrow X' = \uparrow(X' \circ Y') \subseteq \uparrow Y'$. Since $\uparrow X'$ is a maximal face of $\mathcal{M}'$, we get $X' = Y'$. If $\mathrm{Sep}(X', Y') \neq \varnothing$, then by Lemma 10 there exists a geodesic gallery $(\uparrow X' = \uparrow X_0, \uparrow$

$X_1, \dots, \uparrow X_k = \uparrow Y')$ from $\uparrow X'$ to $\uparrow Y'$ in $\mathcal{M}'$. By the definition of a gallery, the union of $\uparrow X'$ and $\uparrow X_1$ is a face $\uparrow Z \supsetneq \uparrow X'$ of $\mathcal{M}'$. Thus, $\uparrow X'$ is not a maximal face of $\mathcal{M}'$ and this contradicts the assumption that X' is a minimal covector of $\mathcal{M}'$. □

4.4. **The labeled compression scheme.** Now, we describe the compression and the reconstruction and prove their correctness. The compression map generalizes the compression map for ample classes of [56]. However, the reconstruction map is much more involved than the reconstruction map for ample classes.

Compression. Let $\mathcal{M} = (U, \mathcal{L})$ be a COM of VC-dimension d. For a sample $S \in \downarrow \mathcal{L}$ of $\mathcal{M}$, consider the tope $S' = S \setminus S^0$ of $\mathcal{M} \setminus S^0 =: \mathcal{M}'$ and let X' be a minimal covector of $\mathcal{M}'$ such that $S' \geq X'$. Denote by $\mathcal{M}'(X') = \uparrow X' \setminus \underline{X'}$ the OM defined by the face $\uparrow X'$ of $\mathcal{M}'$. Define $\alpha(S)_e = S_e$ if $e \in f_{\mathcal{M}'(X')}(S')$ and $\alpha(S)_e = 0$ otherwise. The map α is well-defined since S' is a tope of $\mathcal{M}'(X')$ and hence the sample S' is full in $\mathcal{M}'$. Moreover, by definition we have $\alpha(S) \leq S$, whence $\alpha(S) \in \downarrow \mathcal{L}$. Finally, by Lemma 4 the OM $\mathcal{M}'(X')$ has VC-dimension at most d and thus, by Lemma 14 $\alpha(S)$ has support of size $\leq d$.

Reconstruction. To define β, pick $C \in \{\pm 1, 0\}^U$ in the image of α and let $D := \underline{C}$. Let X be any covector from $\mathcal{H}_D$. By Lemma 15, X exists. Let $\widetilde{S} \in \downarrow \mathcal{L}$ be a sample satisfying:

(1) $\widetilde{S} \geq X$;
(2) $\mathrm{Sep}(\widetilde{S}, C) = \varnothing$;
(3) $\widetilde{S}$ is full in $\mathcal{M}(X)$;
(4) $f_{\mathcal{M}(X)}(\widetilde{S}) = D$.

Finally, set $\beta(C)$ to be any tope T of $\mathcal{M}$ with $T \geq \widetilde{S}$.

To show that β is well-defined we give a canonical sample $\widehat{S}$ satisfying (1)-(4). For $S \in \downarrow \mathcal{L}$, let $C = \alpha(S)$, $D = \underline{C}$, and $X \in \mathcal{H}_D$. By Lemma 15, X satisfies $X \setminus S^0 = X'$, where X' is the minimal covector of $\mathcal{M}' = \mathcal{M} \setminus S^0$ chosen in the definition of $\alpha(S)$. Set $\widehat{S} := X \circ S \geq X$.

Claim 2. $\widehat{S}$ satisfies the conditions (1)-(4) of the definition of β. Moreover, $[\widehat{S}] = [X] \cap [S]$ and $\widehat{T} \geq S$ for any tope $\widehat{T} \in [\widehat{S}]$.

Proof. Let $C = \alpha(S)$ for $S \in \downarrow \mathcal{L}$. Since $X \setminus S^0 = X' \leq S' = S \setminus S^0$, we have $\mathrm{Sep}(X, S) = \varnothing$. By Lemma 12 $[\widehat{S}] = [X] \cap [S]$ is a proper convex subgraph of $[X]$. Since $X \setminus \widehat{S}^0 = X'$ and both $\mathcal{M}(X), \mathcal{M}'(X')$ have the same VC-dimension $|D|$, the sample $\widehat{S}$ is full in $\mathcal{M}(X)$. Since $[\widehat{S}] = [X] \cap [S]$, $\mathrm{osc}([\widehat{S}]) \subseteq \mathrm{osc}([X]) \cup \mathrm{osc}([S])$. Since $S^0 \cap \mathrm{osc}([S]) = \varnothing$, no element of S^0 is osculating with $[\widehat{S}]$ in $[X]$. Thus X' and S' are obtained from X and $\widehat{S}$ by deletion of non-osculating elements of S^0. By Lemma 14(iii) we get $f_{\mathcal{M}(X)}(\widehat{S}) = f_{\mathcal{M}'(X')}(S') = D$. Since $[\widehat{S}] = [X] \cap [S] \neq \varnothing$, this intersection contains at least one tope and therefore $\beta(C)$ is well-defined. Moreover, for any tope $\widehat{T} \in [\widehat{S}]$ we have $\widehat{T} \geq S$ because $[\widehat{S}] = [X] \cap [S]$. □

Correction. We prove that (α, β) defines a proper labeled sample compression scheme, namely, we show that all samples $S \in \downarrow \mathcal{L}$, $\beta(\alpha(S)) \geq S$. We show that for any choice of $\widetilde{S}$ satisfying the conditions (1)-(4) in the definition of β and for any choice of a tope $\widetilde{T} \in [\widetilde{S}]$, we have $\widetilde{T} \geq S$. For this we show that $[\widetilde{S}] = [\widehat{S}]$, where $\widehat{S}$ is the sample defined in Claim 2. Since this implies that $\widetilde{T} \in [\widehat{S}]$, by the second assertion of Claim 2 we get $\widetilde{T} \geq S$. So, let $\widetilde{S}, \widetilde{S}'$ be two full samples of $\mathcal{M}(X)$ such that $\widetilde{S}, \widetilde{S}' \geq X$, $\mathrm{Sep}(\widetilde{S}, C) = \mathrm{Sep}(\widetilde{S}', C) = \varnothing$, and $f_{\mathcal{M}(X)}(\widetilde{S}) = f_{\mathcal{M}(X)}(\widetilde{S}') = D$. By Lemma 14 applied to $\mathcal{M}(X)$, we have $[\widetilde{S}] = [\widetilde{S}']$. This proves that all samples $\widetilde{S}$ available to β yield the same convex subgraph $[\widetilde{S}] = [\widehat{S}] = [X] \cap [S]$. Since $\widehat{T} \geq S$ for any tope $\widehat{T}$ from $[\widehat{S}]$, we conclude that for any choice of $\widetilde{T} \in [\widetilde{S}]$ as $\beta(\alpha(S))$ we have $\widetilde{T} \geq S$. Hence, $\beta(\alpha(S)) \geq S$. This concludes the proof of Theorem 3, the main result of the paper.

Example 1. Consider the tope graph G of a COM $\mathcal{M}$ of VC-dimension 3 and a realizable sample $S = (+ + -0 - 0 + 0)$ in Figure 5. $[S]$ is induced by 7 topes drawn as white vertices of G.

Contracting the 3 dashed Θ-classes corresponding to $\{4,6,8\} = S^0$, yields the tope graph G' of $\mathcal{M}' = \mathcal{M} \setminus S^0$. Then $S' = S \setminus S^0 = (++--+)$. The compressor picks a minimal covector $X' = (+0--+)$ of $\mathcal{M}'$ with $S' \geq X'$; X' corresponds to the thick red edge in G'. The compressor returns $\alpha(S) = (0+000000)$ and $D = \{2\}$. The reconstructor receives $C = (0+000000) = \alpha(S)$, defines $D = \underline{C} = \{2\}$ and constructs the set $\mathcal{H}_D$. There are six covectors of $\mathcal{M}$ belonging to $\mathcal{H}_D$ corresponding to the thick red edges in G. By the localization lemma, they are the covectors which have the same VC-dimension as X' and agree with X' on $\{1,2,3,5,7\} = \underline{S}$. The reconstructor picks $X = (+0----+-) \in \mathcal{H}_D$. The OM $\mathcal{M}(X)$ is composed of the covectors X and the ends T and T' of the corresponding red edge. Among T and T', only the tope $T = (++----+-)$ satisfies the conditions (1)-(4) in the definition of β. Thus, $\beta(\alpha(S))$ is set to T, which is a white node of G.

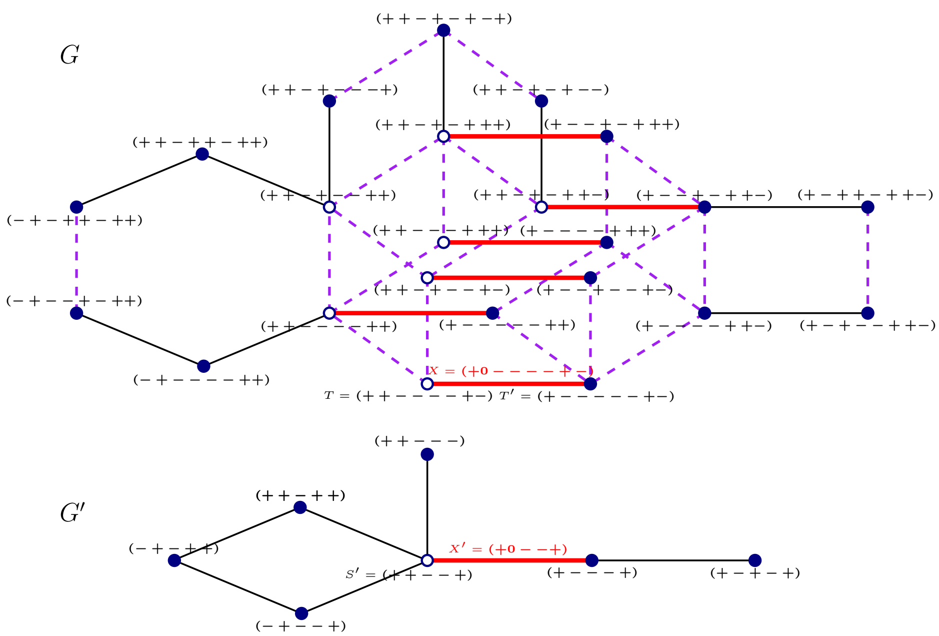

Figure 5. An illustration of Example 1.

5. Conclusion

We have presented proper labeled compression schemes of size d for COMs of VC-dimension d. This is a generalization of the results of [56] for ample set systems, of [7] about affine arrangements of hyperplanes, and of our result [19] about complexes of uniform oriented matroids. Even though we made strong use of the structure of COMs, it is tempting to extend our approach to other classes, e.g., bouquets of oriented matroids [22], strong elimination systems [4], or CW left-regular-bands [53]. Our treatment of realizable samples as convex subgraphs suggests an angle at general partial cubes.

To achieve *improper* labeled compression schemes our results provide a new approach, extending the one of [18, 19] presented in the introduction. Is it possible to extend a given set system or a partial cube to a COM without increasing the VC-dimension too much?

In unlabeled sample compression schemes, the compressor α is less expressive since its image is in 2^U and has to satisfy $\alpha(S) \subseteq \underline{S}$. Unlabeled compression schemes exist for realizable affine oriented matroids [7] and ample set systems with corner peelings [11, 48]. Very recently, Marc [52] designed unlabeled sample compression schemes for OMs. His construction uses Oriented Matroid

Programming and Lemma 14. Moreover, he shows that the recent notion of corner peeling for COMs [47], yields unlabeled compression schemes in this more general setting.

References

[1] M. Albenque and K. Knauer, Convexity in partial cubes: the hull number, Discr. Math. 339 (2016), 866–876.

[2] H.-J. Bandelt, Graphs with intrinsic S_3 convexities, J. Graph Th. 13 (1989), 215—238.

[3] H.-J. Bandelt, V. Chepoi, A. Dress, and J. Koolen, Combinatorics of lopsided sets, Eur. J. Comb. 27 (2006), 669–689.

[4] H.-J. Bandelt, V. Chepoi, and K. Knauer, COMs: complexes of oriented matroids, J. Comb. Th., Ser. A 156 (2018), 195–237.

[5] P.L. Bartlett, N. Harvey, C. Liaw, and A. Mehrabian, Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks, J. Mach. Learn. Res. 20 (2019), 1–17.

[6] A. Baum and Y. Zhu, The axiomatization of affine oriented matroids reassessed, J. Geom. 109, 11 (2018).

[7] S. Ben-David and A. Litman, Combinatorial variability of Vapnik-Chervonenkis classes with applications to sample compression schemes, Discr. Appl. Math. 86 (1998), 3–25.

[8] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. Ziegler, Oriented Matroids, Encyclopedia of Mathematics and its Applications, vol. 46, Cambridge University Press, 1993.

[9] R. Bland, M. Las Vergnas, Orientability of matroids, J. Comb. Th. Ser. B, 23 (1978), 94–123.

[10] B. Bollobás and A.J. Radcliffe, Defect Sauer results, J. Comb. Th. Ser. A 72 (1995), 189–208.

[11] J. Chalopin, V. Chepoi, S. Moran, and M.K. Warmuth, Unlabeled sample compression schemes and corner peelings for ample and maximum classes, J. Comput. Syst. Sci. 127 (2022), 1–28.

[12] J. Chalopin, V. Chepoi, F. Mc Inerney, S. Ratel, and Y. Vaxès, Sample compression schemes for balls in graphs (submitted).

[13] V. Chepoi, d-Convex sets in graphs, Dissertation, Moldova State Univ., Chişinău, 1986.

[14] V. Chepoi, Isometric subgraphs of Hamming graphs and d-convexity, Cybernetics 24 (1988), 6–10.

[15] V. Chepoi, Separation of two convex sets in convexity structures, J. Geometry 50 (1994), 30–51.

[16] V. Chepoi, Distance-preserving subgraphs of Johnson graphs, Combinatorica 37 (2017), 1039–1055.

[17] V. Chepoi, K. Knauer, and T. Marc, Hypercellular graphs: Partial cubes without Q_3^- as partial cube minor, Discr. Math. 343 (2020), 111678.

[18] V. Chepoi, K. Knauer, and M. Philibert, Two-dimensional partial cubes, Electron. J. Comb., 27 (2020), p3.29.

[19] V. Chepoi, K. Knauer, and M. Philibert, Ample completions of Oriented Matroids and Complexes of Uniform Oriented Matroids, SIAM J. Discrete Mathematics, 36 (2022), 505–535.

[20] I. da Silva, Axioms for maximal vectors of an oriented matroid: a combinatorial characterization of the regions determined by an arrangement of pseudohyperplanes, Eur. J. Comb., 16 (1995), 125–145.

[21] P. Deligne, Les immeubles des groupes de tresses généralisés, Inventiones Mathematicae 17 (1972), 273–302.

[22] M. Deza, K. Fukuda, On bouquets of matroids and orientations, Kokyuroku 587 (1986), 110–129.

[23] E. Delucchi and K. Knauer, Finitary affine oriented matroids. arXiv:2011.13348, (2020).

[24] R. DeVore, B. Hanin, and G. Petrova, Neural network approximation, Acta Numerica, 30 (2021), 327–444.

[25] D.Ž. Djoković, Distance–preserving subgraphs of hypercubes, J. Comb. Th. Ser. B 14 (1973), 263–267.

[26] G. Dorpalen-Barry, J. S. Kim, V. Reiner, Whitney numbers for poset cones. Order (2021), 1–40.

[27] A.W.M. Dress, Towards a theory of holistic clustering, DIMACS Ser. Discr. math. Theoret. Comput.Sci., 37 Amer. Math. Soc. 1997, pp. 271–289.

[28] A.W.M. Dress, R. Scharlau, Gated sets in metric spaces, Aequationes Math. 34 (1987), 112–120.

[29] P.H. Edelman, R.E. Jamison, The theory of convex geometries, Geom. Dedicata 19 (1985), 247–270.

[30] J. Edmonds and A. Mandel, Topology of Oriented Matroids, PhD Thesis of A. Mandel, University of Waterloo, 1982, 333 pages.

[31] D. Eppstein, Recognizing partial cubes in quadratic time, SODA 2008, pp. 1258–1266.

[32] D. Eppstein. Isometric diamond subgraphs. In International Symposium on Graph Drawing, pp. 384-389. Springer, Berlin, Heidelberg, 2008.

[33] D. Eppstein, J.-C. Falmagne, and S. Ovchinnikov. Media theory: interdisciplinary applied mathematics. Springer Science & Business Media, 2007.

[34] S. Floyd, On space bounded learning and the Vapnik-Chervonenkis dimension, PhD Thesis, International Computer Science Institut, Berkeley, CA, 1989.

[35] S. Floyd and M.K. Warmuth, Sample compression, learnability, and the Vapnik-Chervonenkis dimension, Machine Learning 21 (1995), 269–304.

[36] J. Folkman and J. Lawrence, Oriented matroids, J. Comb. Th. Ser. B, 25 (1978), 199–236.

[37] B. Gärtner and E. Welzl, Vapnik-Chervonenkis dimension and (pseudo-)hyperplane arrangements, Discr. Comput. Geom. 12 (1994), 399–432.

[38] R. L. Graham, and H. O. Pollak. On the addressing problem for loop switching. The Bell system technical journal, 50 (1971), 2495–2519.
[39] J.E. Grisby and K. Lindsey, On transversality of bent hyperplane arrangements and the topological expressiveness of ReLU neural networks, SIAM J. Appl. Algebra and Geometry (to appear).
[40] F. Haglund and F. Paulin, Simplicité de groupes d'automorphismes d'espaces à courbure négative, The Epstein birthday schrift, Geom. Topol. Monogr., 1 (1998), 181–248. Geom. Topol. Publ., Coventry.
[41] B. Hanin and D. Rolnick, Complexity of linear regions in deep networks, ICML 2019, pp. 2596–2604, PMLR 2019.
[42] B. Hanin and D. Rolnick, Deep ReLU networks have surprisingly few activation patterns, NeurIPS 2019, pp. 359–368.
[43] D. Helmbold, R. Sloan, and M.K. Warmuth, Learning nested differences of intersection-closed concept classes, Machine Learning, 5 (1990), 165–196.
[44] W. Hochstättler and V. Welker, The Varchenko determinant for oriented matroids, Math. Z. 293 (2019), 1415–1430.
[45] V. Itskov, A. Kunin, and Z. Rosen. Hyperplane neural codes and the polar complex. In Topological Data Analysis, pp. 343-369. Springer, Cham, 2020.
[46] K. Knauer and T. Marc, On tope graphs of complexes of oriented matroids, Discr. Comput. Geom. 63 (2020), 377–417.
[47] K. Knauer and T. Marc, Corners and simpliciality in oriented matroids and partial cubes, arXiv:2002.11403, (2020).
[48] D. Kuzmin and M.K. Warmuth. Unlabeled compression schemes for maximum classes, Journal of Machine Learning Research 8, no. 9 (2007).
[49] J. Lawrence, Lopsided sets and orthant-intersection of convex sets, Pac. J. Math. 104 (1983), 155–173.
[50] N. Littlestone, M.K. Warmuth, Relating data compression and learnability, Unpublished, 1986.
[51] T. Marc, Mirror graphs: Graph theoretical characterization of reflection arrangements and finite Coxeter groups. European Journal of Combinatorics 63 (2017), 115–123.
[52] T. Marc, Unlabeled sample compression schemes for oriented matroids, arXiv:2203.11535, 2022.
[53] S. Margolis, F. Saliola, and B. Steinberg, Cell complexes, poset topology and the representation theory of algebras arising in algebraic combinatorics and discrete geometry, Mem. Amer. Math. Soc. 274 (2021), no 1345.
[54] S. Moran, Shattering-Extremal systems, arXiv:1211.2980v1, 2012.
[55] S. Moran, Personal communication, 2021.
[56] S. Moran, M.K. Warmuth, Labeled compression schemes for extremal classes, ALT 2016, 34–49.
[57] S. Moran, A. Yehudayoff, Sample compression schemes for VC classes, J. ACM 63 (2016), 1–21.
[58] T. Mészáros and L. Rónyai, Shattering-extremal set systems of VC dimension at most 2, Electr. J.Comb., 21 (2014), P 4.30.
[59] G.F. Montufar, R. Pascanu, K. Cho, and Y. Bengio, On the number of linear regions of deep neural networks, NIPS 2014, pp. 2924–2932.
[60] T. Neylon, Sparse solutions for linear prediction problems, PhD Thesis, New York Univ., Courant Institut of Mathematical Sciences, 2006.
[61] S. Ovchinnikov, Graphs and Cubes, Springer, New York, 2011.
[62] A. Padrol and E. Philippe, Sweeps, polytopes, oriented matroids, and allowable graphs of permutations. arXiv preprint arXiv:2102.06134 (2021).
[63] D. Pálvölgyi and G. Tardos, Unlabeled compression schemes exceeding the VC-dimension, Discr. Appl. Math. 276 (2020), 102–107.
[64] B.I.P. Rubinstein and J.H. Rubinstein, Geometric and topological representation of maximum classes with application to sample compression, COLT'08, pp. 299–310, 2008.
[65] B.I.P. Rubinstein and J.H. Rubinstein, A geometric approach to sample compressionn, J. Mach. Learn. Res. 13 (2012), 1221–1261.
[66] B.I.P. Rubinstein, J.H. Rubinstein, and P.L. Bartlett, Bounding embeddings of VC classes into maximum classes, in: V. Vovk, H. Papadopoulos, A. Gammerman (Eds.), Measures of Complexity. Festschrift for Alexey Chervonenkis, Springer, 2015, pp. 303–325.
[67] P.W. Shor, Stretchability of pseudolines is NP-hard, The Victor Klee Festschrift, DIMACS Ser. Discr. Math. and Theor. Comput. Sci., 4 (1991), 534–554.
[68] V.N. Vapnik and A.Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, Theory Probab. Appl. 16 (1971), 264–280.
[69] M.K. Warmuth, Compressing to VC dimension many points, COLT/Kernel 2003, 743–744.