with our results of Section 3 implies that the number of samples needed is not more than proportional to $\alpha \log(km)$ where α is $4, 5$ or 6, for Euclidean, similarity and affine transformation, respectively. This is, of course, much smaller than the number of samples required for identification.

Note, however, that the actual VC-dimension of such classes may be even lower than these logarithmic bounds. This is the case for, e.g., the class of translations of convex polygonal objects: Even though the number of sides may be arbitrarily large, the VC-dimension is only three [PW90]. We conjecture here that the true VC-dimension of more general classes of transformed objects will be indeed logarithmic in the complexity of the object.

We assumed that the data available consists of random points inside the object. Computer Vision application usually starts with a much richer set of measurements including boundary slope, curvature, etc. In the full version of this paper we show how to generalize the analysis to deal with this kind of data.

The ability to make the good hypotheses that achieves a small prediction error may be tied to the ability to localize the object within certain precision, by constracting the following mapping between the localization imprecision and the maximal prediction error associated with it: For any (known or bounded) sampling distribution, and any metric used to measure the localization error, one may consider all pairs of object instances whose localization error is above a certain threshold d_0, and find a lower bound on the associated distribution-weighted symmetric differences. A successful learning procedure that results in prediction error smaller than this bound implies that the error in the location of the object is smaller than d_0. Such a procedure was suggested and used in [L92] to set bounds on the probability of achieving several recognition tasks for specific objects.

6 Acknowledgments

We thank Ilan Newman for insightful discussions concerning this work.

References

[BI88] Benedek, G.M., and A. Itai, "Learnability by Fixed Distributions", *Proc. of COLT88*, pp. 80-90. To appear in *Theoretical Computer Science.*

[BEHW89] Blumer, A., A. Ehrenfeucht, D. Haussler and M.K. Warmuth, 1989, "Learnability and The Vapnik-Chervonenkis Dimension", *JACM*, **36**(4), pp. 929-965.

[C65] Cover, T.M., "Geometrical and Statistical Properties of Systems of Linear Equations with Applications in Pattern Recognition", 1965, *IEEE Trans. Electron. Comput.*, pp. 326-334.

[D84] Dudley, R.M., "a course on empirical processes", *Lecture Notes in Mathematics*, 1097, pp. 2-142.

[DCS92] Keren, D., D.Cooper, and J. Subrahmonia, 1992, "Describing Complicated Objects by Implicit Polynomials", accepted to *IEEE Transaction on Patt. Anal. Mach. Intell.*.

[KT61] Kolmogorov, A.N. and V.M. Tichomirov, 1961, "ϵ-entropy and ϵ-capacity of sets in functional spaces", *Amer. Amth. Soc. Translations (Ser. 2)*, 17, pp. 277-364.

[L92] M. Lindenbaum, "Bounds on Shape Recognition Performance", CIS Report 9215, September 1992.

[M64] Milnor, J., 1964, "On the Betti Numbers of Real Varieties", *Proc. Amer. Math. Soc.* **15**, pp. 275-280.

[PW90] Pach, J. and G. Woeginger, 1990, "Some New Bounds on ϵ-Nets ", *Proc. 6th Int. Symp. on Comp. Geometry,* pp. 10-15.

[S72] Sauer, N., 1972, "On the Density of Family of Sets", *Journal of Combinatorial Theory (Series A)*, 13, pp. 145-147.

[TY87] Tannenbaum, A., and Y. Yomdin, 1987, "Robotic Manipulators and the Geometry of Real Semialgebraic sets", *IEEE Journal on Rob. Aut.*, RA-3, pp. 301-307.

[TPBF87] Terzopoulos, D., J. Platt, A. Barr and K. Fleischer, 1987, "Elastically Deformable Models", *ACM Computer Graphics 21*, pp. 205-214.

metric d_U), and there are $C(1/\epsilon, k)$ such distinct unions. □

Repeating the proof of the previous claim, replacing the sets of circles $\mathcal{B}_\epsilon$ by the sets of rectangles $\mathcal{R}_\epsilon \stackrel{def}{=} \{\{(x,y) : \frac{i}{\epsilon} < x < \frac{i+1}{\epsilon}\} : i < 1/\epsilon\}$, one readily gets:

Claim 4.5 *For every* $0 < \epsilon < 1$, $\mathcal{M}_U(\epsilon, SA^2_{(2k,1)}) \geq C(1/\epsilon, k)$.

Using a somewhat more complicated construction this can be improved to:

Claim 4.6 *For every* $0 < \epsilon < 1$, $\mathcal{M}_U(\epsilon, SA^2_{(k,1)}) \geq C(1/2\epsilon, k)$.

Proof: For a given ϵ, let θ be a function, $\theta : \{2i\epsilon : 0 \leq i \leq 1/2\epsilon\} \rightarrow \{0,1\}$, and let p_θ be the polygonal object obtained by successively connecting the points $(i/2\epsilon, \theta(i/2\epsilon))$. Define a class $\mathcal{P}_{(\epsilon,k)} \stackrel{def}{=} \{\{(x,y) : y < p_\theta(x)\} : \theta$ is as above and gets the value 1 no more than k - many times.

Note that, for every ϵ and k there exist $C(1/2\epsilon, k)$ - many such functions θ and each class $\mathcal{P}_{(\epsilon,k)}$ is ϵ-separated. □

The next claim is the reason we bothered with the previous one. It uses the polygonal objects constructed in the previous proof to lower bound the ϵ-entropy of the classes of polynomial objects:

Claim 4.7 *For every integer d and every* $1/d < \epsilon < 1$, $\mathcal{M}_U(\epsilon, SA^2_{(1,d)}) \geq C(1/3\epsilon, d)$.

Proof:[sketch] Given any d and ϵ, we repeat the construction of the classes $\mathcal{P}_{(\epsilon,d)}$ but replace the polygonal objects p_θ by their degree - d Bernstein polynomial approximations. □

The significance of these results to our discussion stems from their comparison to the upper bounds of Section 3 on the VC-dimensions of corresponding classes of transformed objects. Consider a student who is trying to learn by viewing labeled examples drawn independently according to the uniform distribution on the unit square. We compare the information complexity (i.e., the number of labeled examples needed) of two types of tasks: the task of *identification*, in which all the student knows are the parameters (k, m) of the semi algebraic class to which the target object belongs, and the task of *localization* in which he knows that the target is a transformed image of some given object V.

Combining the results of this section with the lower bound of Theorem 4.2, we get the following lower bounds on the information complexity of the task of identification:

For $C = SA^2_{(k,2)}$ or $SA^2_{(k,1)}$ and, when $\epsilon > 1/k$, also for $C = SA^2_{(1,k)}$,

$$l^U_C(\epsilon, \delta) \geq k \log(1/\epsilon) + \log(1-\delta).$$

On the other hand, for the task of localization of any object $V \in SA^2_{(k,m)}$ under the class of all affine transformations, the basic distribution-free upper bound of Blumer et al [BEHW89], combined with our results of Section 3 yields:

$$l^U_C(\epsilon, \delta) \leq \max\{\log(km)\frac{48}{\epsilon}\log\frac{13}{\epsilon}, \frac{4}{\epsilon}\log\frac{2}{\delta}\}.$$

5 The implications on the shape recognition problem.

Consider a student who is trying to learn by viewing labeled examples drawn independently according to some distribution on the unit square. We compare the information complexity (i.e., the number of labeled examples needed) of two types of tasks: the task of *identification*, in which all the student knows are the parameters (k, m) of the semi algebraic class to which the target object belongs, and the task of *recognition* in which he knows that the target is a transformed image of some given object V.

Combining the results of section 2 with the lower bound of Blumer et al [BEHW89], implies that for some 'unfortunate' sampling distribution the student will need a large, number of samples, proportional to km^2, to identify a semi-algebraic object of degree (k, m) up to a prediction error of ϵ.

The results of the last section strengthen this conclusion and imply that this difficulty is not due to some 'peculiar' distribution but exists also when the sampling distribution is uniform. Our results, combined with Theorem 4.2, imply that the number of samples needed for identification is at least proportional to k. We conjecture here that this bound is not tight and that identification of semi algebraic sets under the uniform distribution is as hard as identification under arbitrary distribution.

On the other hand, for the task of localization of any object $V \in SA^2_{(k,m)}$ under the classes of affine, Euclidean and Similarity transformations, the basic distribution-free upper bound of Blumer et al [BEHW89], combined

Where $\rho = min(\epsilon, \delta)$.

The above result reduces the assessment of the information- complexity of the learnability of a class under a given distribution P to finding its covering numbers under d_P. A fundamental result of Dudley now brings us back to the VC-dimension.

Theorem 4.3 (Dudley [D84]) *For any measurable space $(X, \mathcal{O})$, any class $B \subseteq \mathcal{O}$, and any finite subset A,*

$$\inf\{w : \|\{b \cap A : b \in B\}\| \leq \|A\|^w\} = \inf\{w : \sup\{\mathcal{N}_P(\epsilon, B)\} = O(\epsilon^{-w})\},$$

where the suprimum is taken over all probability distributions on $\mathcal{O}$.

By Sauer Lemma [S72], the theorem implies that, for a class C having dimension d, $(1/\epsilon)^d$ is an upper bound on its ϵ-covering numbers relative to any distribution, and that there exist distributions that give rise to covering numbers that are arbitrarily close to this function of ϵ.

Wishing to establish lower bounds on the difficulty of learning under some fixed distribution, we shall have to show that for the classes we care about, the covering numbers relative to this distribution approach the upper bound suggested by the theorem.

One should note that if a class B has a finite VC dimension ,d, then, there always exists a probability distribution, P, such that for $\epsilon = 1/d$, $\mathcal{N}_P(\epsilon, B) = 2^d$ (which is of the same order of magnitude as Dudley's upper bound). To establish this claim just pick a set of size d that is shattered by B and let P be the uniform distribution over this set.

The interesting questions are to show that such 'maximum capacity' behavior can be attained for arbitrarily small ϵ's and, maybe more important, to demonstrate such behavior relative to natural distributions.

The first probability measure that comes to one's mind, when considering a bounded region of a Euclidean space, is the uniform distribution over that region. Clearly, many classes in such a space may have much lower capacities than the bounds derived from their VC dimension. For example, classes of finite (and co-finite) sets may have arbitrarily large VC dimension, yet their metric capacity is just 1 (for any $\epsilon > 0$). We wish to set forth the thesis that, as long as the classes under consideration are natural classes of geometric objects, the bounds derived from the combinatorial considerations are indeed matched by the ϵ-capacities under the uniform distribution.

Given the vagueness of the notion of a 'natural class of geometric objects' we settle for demonstrating the above claim through various examples of such classes. In the full version, we propose a rigorous definition of this notion, and conjecture that a general theorem along these lines does holds for that definition.

Let $C(n, d)$ denote $\Sigma_{i=0}^{d}\binom{n}{i}$, Sauer's Lemma [S72] states that, for any class B having VC-dimension d, $C(n, d)$ is an upper bound on the cardinality of $\Pi_B(A)(= \{b \cap A : b \in B\})$ for sets A of cardinality n, and is the minimal such upper bound. Note that $d = \inf\{w : C(n, d) = O(n^w)\}$. It follows that Dudley's bound is established once one shows that, for any class B having dimension d and for any $\epsilon > 0$, $\mathcal{M}(\epsilon, B) \geq C(1/\epsilon, d)$.

To prove his theorem, Dudley picks, for every n, a set A of cardinality n for which this bound is attained. He then constructs a probability distribution P that concentrates on these sets and gives each member of such a set equal probability weight (of about $1/|A|$)). Under such a distribution, $\Pi_B(A)$ is ϵ-separated and, therefore, $\mathcal{M}_P(1/|A|, B) = C(|A|, d)$, meeting the upper bound of the theorem.

When one wishes to employ this idea to the uniform distribution, the sets A must be chosen more carefully. Having no control over the distribution, our tool for giving the needed weights to members of A is to make sure that, with every point x in such a set, there is an attached neighborhood U_x such that the members of B that participate in defining $\Pi_B(A)$ do not divide these neighborhoods (i.e., there exists $B' \subseteq B$ such that $\Pi_{B'}(A) = \Pi_B(A)$ and, for every $b \in B'$ and $x \in A$, either $U_x \subseteq b$ or $U_x \cap b = \emptyset$). The probabilities of these neighborhood sets, under the uniform distribution, will now play the role that $P(x)$ plays in Dudley's construction of P.

Let us demostrate the the theme discussed above by applying it to a couple of example classes of subsets of the unit square. Let US denote $[0, 1] \times [0, 1]$ and let U denote the uniform distribution over it.

Claim 4.4 *For every $0 < \epsilon < 1$, $\mathcal{M}_U(\epsilon, SA^2_{(k,2)}) \geq C(1/\epsilon, k)$.*

Proof: For any $s, t \in \mathbb{R}$ and $\rho > 0$, let $B((s, t), \rho)$ denote the circle $\{(x, y) : (x - s)^2 + (y - t)^2 < \rho^2\}$. For every $0 < \epsilon < 1$, let $\mathcal{B}_\epsilon$ denote $\{B((i\sqrt{\epsilon} - \frac{\sqrt{\epsilon}}{2}, j\sqrt{\epsilon} - \frac{\sqrt{\epsilon}}{2}), \sqrt{\epsilon}/2) : i, j \leq 1/\sqrt{\epsilon}\}$. $\mathcal{B}_\epsilon$ is a set of cardinality $1/\epsilon$ of disjoint circles, each of which having area $\frac{\pi}{4}\epsilon$. Every union of $\leq k$ of such circles is a member of $SA^2_{(k,4)}$, every pair of distinct such unions is at least $\frac{\pi}{4}\epsilon$ apart (in the change, five-dimensional parameter space is used, with four parameters identical to the Euclidean transformation paramters, and the fifth induced by the weaker constraint $t_1^2 + t_2^2 = t_5^2$.

Now, let $B_{\mathcal{T}}(V), B_{\mathcal{E}}(V), B_{\mathcal{S}}(V)$, and $B_{\mathcal{A}}(V)$ be upper bounds on the VC-dimension of the corresponding transformed object classes when the complexity of the object V, represented by the product km, increases to infinity. Applying the parametrizations introduced above, with the method presented in the proof of theorem 3.6 will yield the required assymptotic bounds. The only difference may be the presence of an additional degree-2 polynomial constraint on the parameters, which is easily incorporated by expressing it as two weak inequality constraints.

Corollary 3.8

$$\begin{aligned} B_{\mathcal{T}}(V) &= 2\log(km) \\ B_{\mathcal{E}}(V) &= 4\log(km) \\ B_{\mathcal{S}}(V) &= 5\log(km) \\ B_{\mathcal{A}}(V) &= 6\log(km) \end{aligned}$$

4 Learnability of Geometric Objects Under the Uniform Distribution

The Vapnik-Chervonenkis dimension of a class can be viewed as a measure of the (information-theoretic) hardness of its distribution-free learnability. I.e., its learnability in a setting where the underlying distribution – the distribution according to which examples are provided and relative to which the accuracy of hypotheses is defined – is unknown to the student and, furthermore, his performance is analyzed in a worst-case setting. Consequently, the Vapnik-Chervonenkis dimension of a class may be readily used to provide upper bounds on the difficulty of learning the class, even in less demanding models of learnability (e.g. when the underlying distribution is known to the student, or is chosen from a limited family of candidate distributions).

The relevance of the VC dimension to hardness (i.e. lower bound) results is not so clear at all. 'Real life' settings are usually much more restricted than what the distribution-free model may reflect. It may very well be the case that a class, whose distribution-free learnability is hard, is easily learnable once the underlying distribution is chosen from among a family of 'realistic' or 'relevant' distributions.

In this section we wish to show that, in the realm of classes of geometric objects in a Euclidean space, this is not the case. I.e., the lower bounds on the difficulty of learnability of such classes, as provided by the Vapnik-Chervonenkis dimension, hold even in the restricted model of one fixed underlying distribution - the uniform probability measure.

The difficulty of learning a concept class under a fixed distribution is best analyzed in the context of metric spaces. Let us begin our discussion by introducing some basic concepts from the theory of metric spaces.

Definition 4 *Let (X, d) be a metric space, let A be a subset of X and $\epsilon > 0$.*

- *$B \subseteq X$ is an* ϵ-cover *for A if for every $a \in A$ there exists some $b \in B$ such that $d(a, b) < \epsilon$.*
- *$\mathcal{N}_d(\epsilon, A)$ is the minimal cardinality of an ϵ-cover for A. (If there is no such finite cover then it is defined to be ∞). $\mathcal{N}_d(\epsilon, A)$ is sometimes refered to as the* ϵ-covering number *of A.*
- *$A \subseteq X$ is* ϵ-separated *if, for any distinct $a, b \in A$, $d(a, b) > \epsilon$.*
- *$\mathcal{M}_d(\epsilon, A)$ is the maximal size of an ϵ-separated subset of A. $\mathcal{M}_d(\epsilon, A)$ is sometimes refered to as the* ϵ-capacity *or the* ϵ-entropy *of A.*

The ϵ- covering numbers and ϵ-capacities are closely related. The following inequalities can be verified (see e.g. [KT61]):

Claim 4.1 *For every metric space (X, d), $A \subseteq X$ and $\epsilon > 0$*

$$\mathcal{M}_d(2\epsilon, A) \le \mathcal{N}_d(\epsilon, A) \le \mathcal{M}_d(\epsilon, A).$$

Given a probability space $(X, \mathcal{O}, P)$, a natural pseudometric, d_P, is induced over the set of measurable sets: For every $a, b \in \mathcal{O}$, $d_P(a, b) = P(a \Delta b)$. We shall sometimes use $\mathcal{N}_P(\epsilon, A)$ to denote $\mathcal{N}_{d_P}(\epsilon, A)$ (and similarly for $\mathcal{M}$).

Benedek and Itai [BI88] investigate learnability with respect to a fixed distributions. The results of Section 4 there imply the following bounds:

Theorem 4.2 (Benedek & Itai [BI88]) *Let $l_C^P(\epsilon, \delta)$ denote the number of random examples needed for (ϵ, δ) - learning of a class C with respect to a probability distribution P. For any probability space $(X, \mathcal{O}, P)$, any concept class $C \subseteq \mathcal{O}$ and any positive ϵ and δ,*

$$\begin{aligned} l_C^P(\epsilon, \delta) &\ge \log(1 - \delta) + \log \mathcal{M}_P(2\epsilon, C) \\ l_C^P(\epsilon, \delta) &\le \frac{54}{\rho}(\ln 1/\rho + \ln \mathcal{N}_P(\rho/2, C)). \end{aligned}$$

will be in the original (nontransformed) semi-algebraic set V. The expression for $\bar{y}$ is a linear function of the parameters, and inserting it into the polynomials $\{f_j\}$ that specify V, induces polynomial sets of equal degree in the parameter space $\{\bar{t}|f_j(\bar{t},\bar{x}) > 0\}$. The union and intersection operations on the polynomial sets of $\mathbb{R}^n$ transform into similar operations on the polynomial sets of the parameter space, and therefore, the set $K_{\bar{x}}^V$ is a semi-algebraic set of degree (k,m) itself. □

For our general result we shall employ (a small modification of) the classical result of Milnor [M64], regarding the number of connected components of polynomial sets.

Lemma 3.5 (A modification of Milnor [M64]) *Let $X \subset \mathbb{R}^n$ is specified by the polynomial inequalities $f_1(x) > 0, \cdots, f_l(x) > 0, f_{l+1}(x) \geq 0, \cdots, f_p(x) \geq 0$ with total degree $d = degf_1 + \cdots + degf_p$. Then $\Psi(X)$, the number of connected components of the set X, satisfies*

$$\Psi(X) \leq \frac{1}{2}(2+d)^n$$

Proof: The original theorem of Milnor provides this relation when all inequalities that specify X are weak (that is, all of them are of the type $f_j(x) \geq 0$). Consider now the sequence of sets $\{X_q\}_{q\in\mathbb{N}}$ specified by the weak inequalities $f_1(x) - 1/q \geq 0, \cdots, f_l(x) - 1/q \geq 0, f_{l+1}(x) \geq 0, \cdots, f_p(x) \geq 0$. Clearly, $X_1 \subset X_2 \subset \cdots \subset X$. and $X = \bigcup_{q\in\mathbb{N}} X_q$. Each of the sets X_q does satisfy Milnor's condition, therefore the number of connected components cannot increase unboundedly with q, implying that any upper bound on the number of connected components, that holds for all the X_q's, holds also for X. □

Let $C_{\mathcal{A}}(V)$ denote the class of all objects obtained by transforming a semi-algebraic object V via an affine transformation.

Theorem 3.6 *For every semi algebraic open set V of degree (k,m) in $\mathbb{R}^n$,*

$$\mathrm{V}Cdim(C_{\mathcal{A}}(V)) = O(n^2 \log km)$$

Proof: Let $S = \{x_1, \ldots, x_N\}$ be a subset of $\mathbb{R}^n$ that is shatterd by the class of transformed objects $C_{\mathcal{A}}(V)$. Let us focus on the parameter space of transformations $\mathbb{R}^{n^2+n}$. The union of boundaries $B_S = \bigcup_{i=1}^{i=N} \partial K_{x_i}^V$ of the semi algebraic open parameter sets $\{K_{x_i}^V\}$ divides the parameter space into connected components. We shall show that the number of such components bounds the number of nonempty sets of the form $W(A,S)$, and therefore the exponent of the size of the shattered set S. We shall then aplly Milnor's Lemma to bound the number of the connected components.

Claim 3.7 *For any pair A, A' of distinct subsets of S, any connected component of $\mathbb{R}^{n^2+n} \setminus B_S$ that has a nonepty intersection with $W(A,S)$ neccessarily has an empty intersection with $W(A',S)$.*

Proof:[of the claim] Let x_{i_0} be a point in $A \setminus A'$. By claim 3.3, any $t \in W(A,S)$ is a member of $K_{x_{i_0}}^V$ while for any $t \in W(A',S)$, $t \notin K_{x_{i_0}}^V$. It follows that any path (in the parameter space) connecting a member of $W(A,S)$ with a member of $W(A',S)$, neccessarily meets $\partial K_{x_{i_0}}^V$. □

To calculate the number of connected components of the paramter space recall that each of the parameter sets $K_{x_i}^V$ was specified by k-many polynomial sets of the form $\{\bar{t}|f_j(\bar{t},x_i) > 0\}$ and note that at least one of the functions $f_{ij}(\bar{t}) = f_j(\bar{t},x_i)$ vanishes on each point of the boundary of V_{x_i}.

Consider now the product function $G(t) = \prod_{i,j} f_{ij}(t)$. Any connected component of $\mathbb{R}^{n^2+n} \setminus B_S$ is a union of one or more connected components of $\{t : G(t) > 0\}$ or of $\{t : G(t) < 0\}$. $G(\bar{t})$ is a (kmN)-degree polynomial in n^2+n real variables, and, by our modification to Milnor theorem, the number of connected components of its positive set $\{P|G(\bar{t}) > 0\}$ (as well as of its negative set, which is the pos-set of $-G$) is not higher than $(2+kmN)^{n^2+n}$. The theorem now follows by a straightforward calculation. □

In Computer Vision and Robotics context, one is usually interested only in two or three dimensional object spaces. For these cases, it is easy to see that subgroups of affine transformations, such as Translation, Euclidean and Similarity transformations can be represented in a parameter space smaller than $\mathbb{R}^{n^2+n}$, implying that the corresponding classes of transformed objects $C_{\mathcal{T}}(V)$, $C_{\mathcal{E}}(V)$, $C_{\mathcal{S}}(V)$, have lower VC-dimensions.

In the two dimensional case, for example, the translation transformation may be represented only by the two components of the (inverse) translation vector b'. To represent the Euclidean transformation we need also to specify the rotation matrix A'. To keep the transformation linear in the parameters, which is essential for the derivation, we use a four dimensional space $\bar{t} = \{t_1 = a'_{11}(= a'_{22}), t_2 = a'_{21}(= -a'_{12}), t_3 = b'_1, t_4 = b'_2\}$ with a constraint $t_1^2 + t_2^2 = 1$ kept in mind. For Similarity transformations, which also allow uniform scale to some geometric issue in the space of transformations T. Once this translation is established, the mighty tools of algebraic geometry can be called in to solve our combinatorial problem. We shall now describe a general framework for such a translation.

Given a family of transformations of $\mathbb{R}^n$, T, the first step we take is to note that, any 'object' set, $V \subseteq \mathbb{R}^n$, induces a mapping of points of $\mathbb{R}^n$ to subsets of T. Namely, every point $\bar{x} \in \mathbb{R}^n$ is mapped into the subset of transformations $K_{\bar{x}}^V = \{t \in T : \bar{x} \in V_t\}$. Note that this mapping is dual to the mapping from members of T to subsets of $\mathbb{R}^n$ defined by mapping t to the set V_t.

Before we proceed, let us introduce some further notation: Given a set S, for every subset, $A \subseteq S$, let $\theta_A(x_i)$ be the indicator function:

$$\theta_A(x_i) = \begin{cases} 1 & x_i \in A \\ -1 & else \end{cases}$$

Let us also use the exponent notation R^1 and R^{-1} to denote a subset, $R \subseteq T$, and its complement $T \setminus R$, respectively.

Consider now the set of points $S = \{x_1, x_2, \cdots x_N\}$ in $\mathbb{R}^n$. Fixing an object set $V \subseteq \mathbb{R}^n$ every subset $A \subset S$ corresponds to a subset of the transformations set T

$$W(A, S) = \bigcap_{i=1}^{N} K_{x_i}^{\theta_A(x_i)}$$

(In an attept to simplify the notation, we have dropped the superscript V from the sets K_x^V). The following claim is straightforward from the definitions:

Claim 3.3 *For any $A \subseteq S$ and $t \in T$,*

- *$A = V_t \cap S$ iff $t \in W(A, S)$.*
- *For $t \in W(S, A)$ and $x \in S$, $t \in K_x^V$ iff $x \in A$.*
- *For any object $V \subseteq \mathbb{R}^n$ and a family T of transformations, the class of transformed objects, $C_T(V)$, shatters a set $S \subseteq \mathbb{R}^n$ iff, neither of the members of $\{W(A, S) : A \subseteq S\}$ is empty* [2]

We have therefore reduced the calculation of the VC-dimension of classes of transformed images to counting the number of non-empty $W(A, S)$ sets of transformations. We shall apply this reduction in our subsequent derivations.

[2] Dudley [D84] defines a collection of sets $\{A_1, A_2, \ldots, A_N\}$ to be *independent*, if for every function $\theta : \{1, \ldots, N\} \to \{-1, 1\}$, the intersection, $\bigcap_{i=1}^{N} A_i^{\theta(i)}$, is nonempty. With this notation the above claim says that the class of transformed objects, $C_T(V)$, shatters a set $S \subseteq \mathbb{R}^n$ iff $\{K_{x_i}^V | x_i \in S\}$ is an independent set.

Parametrizing the transformations

The next step we take is to represent T parametrically, with parameters forming some parameter space $\mathbb{R}^k$. For example, if T is the family of translations of $\mathbb{R}^n$, it can be naturally characterized by the n orthogonal translations, as done in the proof of Theorem 3.2 above.

As mentioned in the introduction, the transformations of interest are Translations, Euclidean, Similarity and Affine transformations. All these families are special cases of linear affine transformations. A linear affine transformation on $\mathbb{R}^n$ is defined by a pair, (A, b), where A is an $n \times n$ matrix and $b \in \mathbb{R}^n$. Such a transformation $H = (A, b)$ acts on $\mathbb{R}^n$ by $H(x) = Ax + b$. We shall restrict our attention to non-singular transformations, i.e., transformations that are one-to-one or, equivalently, their defining matrix A is regular. For these transformations the inverse transformation, denoted $H' = (A', b')$, always exists, and its parameters, the components of A' and b' will be used to represent the transformation. Clearly, the parameter vector, denoted $\bar{t}$, is a point in the affine parameter space $\mathbb{R}^{n^2+n}$. Now $K_{\bar{x}}^V$ will also denote the set of parameters that correspond to all transformations t for which V_t includes the point $\bar{x}$. Other transformation subsets, such as the $W(A, S)$ subsets will also denote parameter subsets.

The families of Euclidean transformations and of similarity transformations of $\mathbb{R}^n$ can always be represented in $\mathbb{R}^{n^2+n}$, but can also be represented in lower dimensional parameter spaces. This will enable us to obtain, in some cases, better bounds on the VC-dimension of classes of images of an object under such transformations.

The following Lemma is the main tool for the passage from the object space to the space of transformations. It extends, to arbitrary semi algebraic sets and to arbitrary affine transformatios, a property that was already used for translations of polynomial sets in Claim 3.2.

Lemma 3.4 *For any semi algebraic set $V \subseteq \mathbb{R}^n$ of degree (k, m) and for every $\bar{x} \in \mathbb{R}^n$, the set of transformation parameters $K_{\bar{x}}^V$ is also a semi-algebraic set of degree (k, m) (in the parameter space of affine transformations $\mathbb{R}^{n^2+n}$).*

Proof: A sufficient and neccessary condition for a point $\bar{x}$ to be inside the transformed object is that the result of applying the inverse transformation on it, $\bar{y} = A'\bar{x} + b'$

Example 1 *Consider the class of convex k-gon. This class is a subclass of $SA^2_{(k,1)}$ and its VC-dimension is at least $2k$.*

(This can be easily proved by looking at a set of $2k$ points equally spaced on the boundary of a circle and noting that, for every subset of it, there is a convex k-gon that contains this subset and no other point of the set).

3 The VC-dimension of classes of transformed objects

In this section we wish to bound from above the VC-dimension of classes of transformed objects. As the Claim 3.1 below demonstrates, such a bound depends on the object generating the class. We shall consider semi-algebraic open sets of a given degree. We will show that, for wide classes of transformations, the dimension of the corresponding classses of transformed images of any (k,m)-semi-algebraic planar set is substantially below the dimension of $SA^n_{(k,m)}$ - the class of all sets of the same degree.

The families of transformation we consider include the families of translations, rotations, scale changes, combinations of them, and the group of all affine transformations.

Definition 3 *Let V be a subset of $\mathbb{R}^n$ and let T be a family of transformations of $\mathbb{R}^n$.*

- *For any $t \in T$, let V_t denote the t-transformed image of V. I.e., $V_t = \{t(\bar{x}) : \bar{x} \in V\}$.*
- *The class of T-transformed images of V, denoted by $C_T(V)$, is $\{V_t : t \in T\}$.*
- *A* translation *is a transformation of $\mathbb{R}^n$ defined for a vector $\bar{t} \in \mathbb{R}^n$ by $t(\bar{x}) = \bar{x} + \bar{t}$.*

Claim 3.1 *There exists a subset of the real line such that the class of its translations has an infinite VC-dimension.*

Proof: Consider the set $S = \{1/2, 1/3, 1/4, \cdots\}$. Let $\{A_i\}_{i\in\mathbb{N}}$ enumerate all its finite subsets. Consider now the set $\mathring{A} = \bigcup_{n\in\mathbb{N}}(A_n + n)$. We claim that the class of its translations, $C_T(\mathring{A}) = \{\mathring{A} + t : t \in \mathbb{R}\}$, has an infinite VC-dimension. To see this, just note that $C_T(\mathring{A})$ shatters any finite subset of S. □

The above claim makes it clear that simple transfrmations do not neccessarily elicit simple classes. In the rest of the section we shall show that the complexity of the transformed object classes depends on the complexity of both the object and the family of the transformations. To start with, we consider the family of translations applied to planar polyomial sets. Classes of translations of a polynomial object are easy to analyze, and will serve as an illustration to the general methodology employed later.

Theorem 3.2 *Let V be a polynomial object of degree m in $\mathbb{R}^2$, and let T be the family of translations in $\mathbb{R}^2$. Then,*

$$VCdim(C_T(V)) = O(\log m).$$

Proof: Let f_V be an m-degree polynomial s.t. $V = \{\bar{x} : f_V(\bar{x}) > 0\}$. Let S be a set of N points shattered by the class $C_T(V)$. That is, every subset of S can be written as $S \cap V_t$ for some $t \in \mathbb{R}^2$. Let K_{x_i} be the set of translation vectors, t, for which a particular point $\bar{x}_i$ of S is included in V_t, $K_{x_i} = \{t : \bar{x}_i \in V_t\}$. Note that $K_{x_i} = \{t : f_V(\bar{x}_i - t) > 0\}$. Therefore, K_i itself is a polynomial set of degree m in $\mathbb{R}^2$.

Let ∂K_{x_i} be the boundary of K_{x_i} and $B_S = \cup_{i=1}^N \partial K_{x_i}$. B_S is the union of closed curves and divides the rest of the plane, $\mathbb{R}^2 \setminus B_S$, into connected components such that all translations in the same connected component correspond to translated objects that contain the same subset of S. Therefore, the number of connected components of $\mathbb{R}^2 \setminus B_S$ is an upper bound to the number of different subsets of S that can be written as $S \cap V_t$.

The number of connected components, $\Phi(N,m)$, induced by N closed degree-m polynomial boundaries can be bounded using some classical results from algebraic geometry, and was investigated before. Tannenbaum and Yomdin [TY87] rely only on the classical theorems of Bezout and Harnack that bound the number of self intersections and mutual intersections of two polynomial curves and show that

$$\Phi(N,m) \leq (2N^2 - N)m^2 - Nm + 2N$$

As the number of subsets of S represented as $S \cap V_t$ is bounded by the number of connected components of $\mathbb{R}^2 \setminus B_S$, it is also bounded by this expression. By our assumption that S is shattered by $C_T(V)$, this should be higher than 2^N. Therefore, we get the relation:

$$2^N \leq (2N^2 - N)m^2 - Nm + 2N \leq 2N^2m^2$$

which limits the size of sets shattered by translations of degree m polynomial object and implies the claim. □

The main idea behind the above proof is to translate the issue of the combinatorial complexity of a class $C_T(V)$

lowing claim is a straight-forward application of this theorem:

Claim 2.2 *The VC-dimension of $SA^n_{(1,m)}$ – the class of all polynomial objects of degree m – is $CC(n,m)$.*

(This claim also follows from the results presented by Cover [C65].) As another application of Theorem 2.1, consider the classes of regions bounded by polynomial functions,

$$B^2_m \stackrel{\text{def}}{=} \{(x,y) \in \mathbb{R}^2 : y < p(x)\} : \quad p \text{ is a polynomial of degree at most } m\}$$

Claim 2.3 *For every $m \in \mathbb{N}$, the VC-dimension of the class B^2_m is $m+1$.*

The proof is immediate by noting that B^2_m equals $\{pos(f) : f(x,y) = p(x) - y\}$ where the degree of p is at most m and y is a fixed function. We shall now present some tools for the calculation of the VC-dimension of classes of $\mathbb{R}^n$ that are generated from classes whose dimensions are known.

Definition 2 *Let C be a class of subsets of $\mathbb{R}^n$,*

- *C, is* shift invariant *if, for every $c \in C$ and for every $\bar{t} \in \mathbb{R}^n$, $c + \bar{t} \stackrel{\text{def}}{=} \{\bar{x} + \bar{t} : \bar{x} \in c\}$ is also in C.*
- *C is* scale invariant *if, for every subset $c \in C$ and for every $\alpha \in \mathbb{R}$, $\alpha c \stackrel{\text{def}}{=} \{\alpha x | x \in c\}$, is also in C.*
- *The* elementwise union *of two classes, $C_1 \underline{\cup} C_2$, is $\{a \cup b | a \in C_1, b \in C_2\}$*
- *The* elementwise intersection *of two classes, $C_1 \underline{\cap} C_2$, is $\{a \cap b | a \in C_1, b \in C_2\}$*

Lemma 2.4 *Let C_1, C_2 be classes of bounded subsets in $\mathbb{R}^n$ whose VC-dimension is finite. Then*

1. *If C_1 and C_2 are shift invariant, or scale invariant, then so are $C_1 \underline{\cup} C_2$ and $C_1 \underline{\cap} C_2$.*
2. *If C is scale invariant, then, for every $r > 0$, bounding C to the ball around the origin $B(o,r) = \{\bar{x} \in \mathbb{R}^n : ||\bar{x}|| \leq r\}$, does not affect its VC-dimension. I.e., $VCdim(C) = VCdim(C \underline{\cap} \{B(0,r)\})$.*
3. *If C_1 and C_2 are shift invariant then $VCdim(C_1 \underline{\cup} C_2) \geq VCdim(C_1) + VCdim(C_2)$.*

Proof: We leave the proof of the first two claims to the reader. For the third claim, Consider two sets of points S_1, S_2, of sizes $VCdim(C_1)$ and $VCdim(C_2)$ respectively, each shattered by the corresponding class. Let SC_1, SC_2 be subclasses of C_1, C_2, that contain the minimal number of subsets needed to shatter these sets. The shift invariance of C_1, C_2 and the boundness of their members, imply that the S_1, S_2 sets can be chosen such that the intersection between any member of SC_1 and any member of SC_2 is null. Every union of an element of SC_1 with an element of SC_2 is a member of $SC_1 \underline{\cup} SC_2$. It follows that the class $SC_1 \underline{\cup} SC_2$, which is a subclass of $C_1 \underline{\cup} C_2$, shatters $S_1 \cup S_2$, implying that its VC-dimension is at least $VCdim(C_1) + VCdim(C_2)$. □

To see how tight this bound is, we compare it to the upper bound derived in [D84]. There, Dudley considers the elementwise union (or intersection) of two classes, C_1 and C_2. He applies Sauer Lemma [S72] to get:

$$VCdim(C_1 \underline{\cup} C_2) \leq sup\{r \in \mathbb{N}; \sum_{i=0}^{VCdim(C_1)} \binom{r}{i} \sum_{j=0}^{VCdim(C_2)} \binom{r}{j} > 2^r\}.$$

Now, letting $r^* = VCdim(C_1 \underline{\cup} C_2)$, Dudley's inequality becomes:

$$\begin{aligned} 2^{r^*} &< \sum_{i=0}^{VCdim(C_1)} \binom{r^*}{i} \sum_{j=0}^{VCdim(C_2)} \binom{r^*}{j} \\ &< (r^*)^{VCdim(C_1)} (r^*)^{VCdim(C_2)} \end{aligned}$$

implying

$$\frac{VCdim(C_1 \underline{\cup} C_2)}{\log(VCdim(C_1 \underline{\cup} C_2))} \leq VCdim(C_1) + VCdim(C_2).$$

We use Lemma 2.4 to obtain a general lower bound on the VC-dimension of semi algebraic sets.

Claim 2.5 *The VC-dimension of the class of semi-algebraic open objects of degree (k,m), is at least $\frac{k}{2} CC(n,m)$.*

Proof: Consider the class $A = B \underline{\cup} C$, where B is the class of polynomial objects of degree m, and C the class of balls (of finite radii). By Claim 2.2 there exists a set of size $CC(n,m)$ that is shattered by the class B. Since there exists some ball in C that contains it, the VC-dimension of the new class A is at least $CC(n,m)$. Note that A contains only bounded sets, and is shift and scale invariant. Applying Lemma 2.4 we conclude that the VC-dimension of the class $\underline{\cup}_{i=1 \cdots k/2} A$ is at least $\frac{k}{2} CC(n,m)$. The claim now follows by noting that this latter class is a class of semi algebraic sets of degree (k,m). □

The following example shows that the above bound is not tight:

transformed classes may be arbitrarily difficult to learn (see section 3.1 for an example). We shall therefore limit our discussion to objects that are semi-algebraic sets in $\mathbb{R}^n$, i.e., can be defined by boolean combinations of polynomial inequalities.

As for the families of allowed transformations, we shall consider affine transformations of $\mathbb{R}^n$ as well as some subgroups such as isometries (or Euclidean transformations), which corespond to repositioning of rigid bodies, and Similarity transformations, which allow also uniform scale change, and models image acquisition distortion commonly present when the information comes through a picture taken by a camera whose distance to the object is unknown.

In Section 2 we define our objects of research - the Semi Algebraic Sets. We then develop some tools for proving lower bounds on the VC-dimension of classes of such sets. Section 3 is the heart of this work - it investigates the VC-dimension of classes of transformed images of a semi algebraic object. We mannage to reduce the problem to a geometric issue in the space of the parameters of the transformations. Using a classical result of Milnor we obtain upper bounds on the VC-dimensions of these classes.

Viewed from the distribtion-free PAC learnability angle, these results imply, on one hand, upper bounds on the number of examples needed for the localization of an object of some semi algebraic degree, and on the other hand, much higher lower bounds on the number of examples needed for the identification of such an object from among all objects of the same degree.

Section 4 carries these results over to the setting of learnability with respect to the (fixed) uniform distribution. This is done by analysing the ϵ-entropy of the relevant classes, and showing that, for wide classes of semi algebraic objects, the entropies under the metric induced by the uniform distributions approach their maximum possible values (over all probability distributions).

Finally, in section 5, several conclusion are drawn regarding the relation of our results to object recognition.

2 The VC-Dimension of classes of Semi-Algebraic sets

We wish to show that localization is, in some sense, an easy task. This statement may fail when the object one wishes to localize is very 'wild', an example of such a case is given later, (see claim 3.1.) We shall therefore focus on well behaved geometrical objects - the Semi-Algebraic subsets of $\mathbb{R}^n$.

Definition 1 • *A Boolean combination of sets is called* positive *if it involves only the union and intersection operations.*

- *A* semi-algebraic open set *of degree (k,m) in $\mathbb{R}^n$ is a set that can be represented as a positive boolean combination of k sets of the form $\{\bar{x} \in \mathbb{R}^n : f_j(\bar{x}) > 0\}$ where the functions f_j are real polynomials of maximal degree m.*
- *A* polynomial set *is a semi-algebraic open set of degree $(1,m)$, for some finite m.*

Notes

- Unlike the common definition of semi algebraic sets (see e.g. [TY87]), our definition of semi algebraic open sets does not allow sets defined by weak inequalities and equalities. This is a technical difference and does not affect the implications of our results to Learnability and Object Recognition. It does, however, help to simplify our proofs.
- From the Computer Vision point of view, even polynomial objects of modest degrees (e.g. 4) seem to enable the description of complicated objects, thereby providing sufficient representation power [DCS92] [TPBF87]. The class we consider here is even richer: besides polynomial objects it also contains combinations of them which include, e.g., polygonal objects (which, for k being the number of polygon sides, are semi algebraic sets of degree $(k,1)$).

We wish to calculate the VC-dimension of classes of the form:

$$SA^n_{(k,m)} \stackrel{\mathrm{def}}{=} \left\{ \begin{array}{l} A \subseteq \mathbb{R}^n : A \quad \textit{is a semi-algebraic} \\ \text{open set of degree } (k,m) \end{array} \right\}$$

The following theorem of Dudley will be useful in the derivations.

Theorem 2.1 (Dudley [D84]) *For a real-valued function f on some domain X, let* $\mathrm{pos}(f)$ *denote $\{x \in X : f(x) > 0\}$. If $\mathcal{H}$ is a real vector space of functions from X ro $\mathbb{R}$ and h is any real-valued function on X then the VC-dimension of $\{\mathrm{pos}(f+h) : f \in \mathcal{H}\}$ equals the linear dimension of $\mathcal{H}$.*

Let $CC(n,m) = \binom{n+m}{m}$ be the number of different monoms in a degree-m polynom of $\mathbb{R}^n$. Then the fol-

Localization vs. Identification of Semi-Algebraic Sets

Shai Ben-David [*] **Michael Lindenbaum** [†]
Department of Computer Science
Technion, Haifa 32000, Israel

Abstract

How difficult is it to find the position of a known object using random samples? We study this question, which is central to Computer Vision and Robotics, in a formal way. We compare the information complexity of two types of tasks: the task of *identification*, in which all the student knows is a description of a natural class to which the object belongs, and the task of *localization* in which he knows that the target is a transformed image of some given object. We model localization as the task of learning the class of transformed instances of the given object. We apply some fundamental results from Algebraic Geometry to bound the VC-dimension of such 'transformed class' and compare it to the VC-dimension of some natural library classes to which the objects belong. We carry on the comparison to the scenario of learning under the uniform distribution, which leads us to calculating the ϵ-entropy of relevant classes. Our analysis provides a mathematical ground to the intuition that Localization is indeed much easier than Identification.

Keywords
Learning Theory, PAC, Vapnik-Chervonenkis dimension, localization, identification, recognition, Computer Vision.

[*]email: shai@cs.technion.ac.il
[†]email: mic@cs.technion.ac.il

1 Introduction

Object recognition, a fundamental task of computer vision, usually deals with the following situation: one observes a scene, extracts some measurements out of it, and uses them to judge whether certain objects are present in the scene, and what are their positions. In the basic form of this task, called localization, the identity of the object is known, and one tries to guess its position correctly. A fundamentally different and more general task is identification, where the object is included in a known library but is otherwise unknown, and its shape as well as its position are to be determined.

Intuitively it seems clear that localization is an easier task than identification. The aim of this work is to provide some mathematical justification to this intuition.

Let a *class of transformed objects* be a class of objects that are transformed instances of one particular object. Such a class depends on the original object and on the type of transformations allowed. One can view localization as the task of identification from a library that is a class of transformed objects [1].

We wish to quantify the 'complexity' of classes of transformed objects, for different objects and groups of transformations, and to compare it to the 'complexity' of some natural library classes to which the objects belong. The measures of complexity of a class that we shall investigate are two: The Vapnik-Chervonenkis dimension and the metric (ϵ-) entropy. These measures are relevant to the learning difficulty of classes in the distribution-free PAC model and with respect to the fixed uniform distribution (repectively).

If no limitations are imposed on the shape of the object,

[1]In the field of computer vision, the term 'recognition', or 'model based recognition' refers to identification from a library that is generated by transforming a finite number of base objects.