# The maximum discrepancy of simple geometric ranges

David P. Dobkin [1] [2] Dimitrios Gunopulos [1] [3]

## Abstract

In this paper we give algorithms that compute the maximum bichromatic discrepancy for simple geometric ranges and consequently solve the minimizing disagreement problem of machine learning for geometric hypotheses. The main result, presented in section 3 is an $O(n^{\min(\alpha+1/2,2k-1)}\log n)$ algorithm that computes the maximum bichromatic discrepancy w.r.t. $k$-gons in two dimensions (where $\alpha$ is the VC-dimension of the implied set system, $k$ is constant). This is an application of a general approach to computing the discrepancy of set systems with finite VC-dimension. We also present an algorithm for the special case of stripes.

## 1. The minimizing disagreement problem

An important goal of applied machine learning is to provide tools for the design and analysis of learning algorithms that provide satisfactory solutions for real-world learning problems. There are a lot of experimental results regarding the performance of various heuristic learning algorithms on a number of "benchmark"-datasets for real world classification problems ([BN], [Ho], [Ma], [WK90], [WGT], [WK91]).

Concept learning is a typical machine learning problem. The objective here is to find a classifier for a concept. We are given a labeled training sequence (a point set of examples, each labeled either as a positive example of the concept or a negative example), and we want to find a hypothesis that can be used as a predictor for other query points ([WK91]). An important optimization problem arises in machine learning, the *minimizing disagreement problem*: Given a labeled training sequence $T$, and a set of hypotheses $\mathcal{H}$, find a hypothesis $H \in \mathcal{H}$ that minimizes the error for the input set. For a given hypothesis $H$, the error of $H$, is the number of points in $T$ that $H$ classifies incorrectly ([Ma]).

[1]The research work of these authors was supported by NSF Grant CCR93-01254 and by DIMACS, an NSF Science and Technology Center.

[2]Department of Computer Science, Princeton University, 35 Olden St., Princeton, NJ 08540, USA, e-mail: dpd@cs.princeton.edu

[3]Department of Computer Science, Princeton University, 35 Olden St., Princeton, NJ 08540, USA, e-mail: dg@cs.princeton.edu

Usually in practice the set of hypotheses $\mathcal{H}$ is the set of hyperplanes. The dimension of the space the input set is drawn from is the number of the attributes of the concept, and a hyperplane is the simplest separator between positive and negative examples. Such linear separators are widely used in practice because they are relatively fast to compute, give intuitive solutions, and degrade well ([WK91]). Their simplicity however often limits their accuracy. There is empirical evidence that more complex models can work particularly well in practice. Weiss and Kulikowski ([WK91]) report that for some real life data sets, hypotheses defined as the disjunction or conjunction of simple rules of the form $x_i \leq c_1$ or $x_j = c_2$ outperform neural nets and decision trees. In [DGM] we present an algorithm to find the optimal hypothesis, when the hypothesis set is restricted to axis aligned boxes.

In this work we examine an extension of the linear separator model. We consider hypotheses that are defined as the intersection of several halfspaces. Hypotheses are thus piecewise linear and convex. The solutions they offer are intuitive and, because the model is more general, more accurate that single halfspaces. We present a polynomial algorithm to solve the minimizing disagreement problem in two and three dimensions for convex hypotheses of constant size.

To formalize the problem, we turn to the concept of the bichromatic discrepancy of set systems ([C93a]): A set system is a pair $(S, \mathcal{R})$, where $S$ is a set of points, and $\mathcal{R}$ is a set of subsets (we will call them ranges) of $S$. Let $(S, \mathcal{R})$ be a set system and let $\chi$ be a mapping of $S$ to the real numbers ($\chi : S \rightarrow \Re$). For a set $R \in \mathcal{R}$, let $\Delta(R) = \sum_{x \in R} \chi(x)$ be the *bichromatic discrepancy* of $R$. We define the *maximum bichromatic discrepancy* of $\chi$ on $(S, \mathcal{R})$ by:

$$\Delta(S, \chi, \mathcal{R}) = \max_{R \in \mathcal{R}} |\Delta(R)|$$

Usually $\chi$ is a mapping to $\{-1, +1\}$, and is called a coloring of $S$. To simplify things and avoid the absolute values, we also define two more functions:

$$\Delta_{max}(S, \chi, \mathcal{R}) = \max_{R \in \mathcal{R}_2} \Delta(R) \quad \text{and} \quad \Delta_{min}(S, \chi, \mathcal{R}) = \min_{R \in \mathcal{R}_2} \Delta(R)$$

Clearly $\Delta(S, \chi, \mathcal{R}) = \max(\Delta_{max}(S, \chi, \mathcal{R}), -\Delta_{min}(S, \chi, \mathcal{R}))$. Since an algorithm that computes $\Delta_{max}(S, \chi, \mathcal{R}$ would also compute -$\Delta_{min}$ if all weights have their sign reversed, we focus on $\Delta_{max}$.

Let $\mathcal{F}_k$ be the set of all planar convex $k$-gons, where a (potentially open) $k$-gon is defined as the intersection of $k$ halfspaces. $k$ is a (small) constant. For given finite set $S \subset [0,1]^2$, the set system $(S, \mathcal{F}_k(S)$ is defined in the following natural way: a subset $A$ of $S$ is in $\mathcal{F}_k(S)$ if and only if there exists a planar convex $k$-gon $W$ that contains exactly $A$ ($S \cap W = A$, Fig. 1).

The minimizing disagreement problem and the problem of computing the maximum discrepancy are equivalent ([DGM]). To solve the minimizing disagreement problem for the set of $k$-gon hypotheses and for a training sequence $T$, we have to find $\Delta_{max}(S, \chi, \mathcal{F}_k(S))$, thet is, the $k$-gon that maximizes $\Delta$ for the set of points $S$ (comprising the points of the

examples in $T$) and the mapping $\chi : S \to \{-1, +1\}$ which is defined by the labels of the examples in $T$.

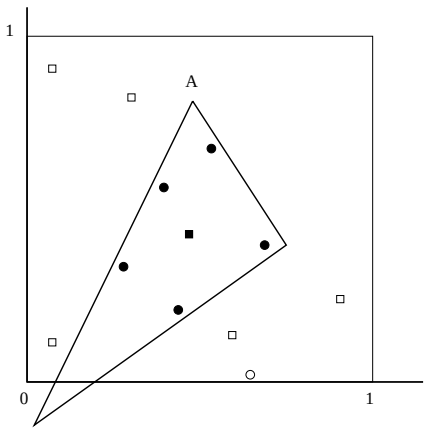


Figure 1: This triangle maximizes the bichromatic discrepancy of this point set for all convex sets (circles are mapped to +1, and squares are mapped to -1, the filled points are the ones in the interior).

In the next two sections we present an efficient algorithm that, given a set $S \in [0,1]^2$ (for the rest of this paper, we assume that $|S| = n$) and a weight function $\chi : S \to \Re$, computes $\Delta_{max}(S, \chi, \mathcal{F}_k(S))$, and therefore solves the equivalent minimizing disagreement problem. To simplify the presentation, we assume that the input points are in general position, that is, there are no three collinear points. If this assumption does not hold, we can use a perturbation technique on the input point coordinates.

The bichromatic discrepancy is the discrete restriction of the numerical discrepancy, which is a measure of the approximation of a continuous function by a point set. In fact, some upper bound results for the numerical discrepancy ([BC]) were obtained via theorems for the bichromatic discrepancy. In addition, the problem of computing the rectangle discrepancy has important applications in the field of computer graphics ([DE], [DM]). Algorithms for the bichromatic discrepancy have been shown to be a good foundation for solving numerical discrepancy case ([DGM]), and, in addition, the numerical discrepancy can be approximated by the substitution of the continuous function by a suitable grid of points.

# 2. The Discrepancy and the VC-dimension of $k$-gons

We begin this section with a closer examination of the set system $(S, \mathcal{F}_k(S))$. Let us first give the definition of the VC-dimension of set systems and related concepts. Let $(S, \mathcal{R})$ be a set system. For a set $Y \subset S$, we call the set system $(Y, \{U \mid (U = R \cap Y) \wedge (R \in \mathcal{R})\})$ the

subspace induced by $Y$. We say that $Y$ is shattered by $\mathcal{R}$ if, in the subspace induced by $Y$, every possible subset of $Y$ is a range (in other words, if $|\{U \mid (U = R \cap Y) \wedge (R \in \mathcal{R})\}| = 2^{|Y|}$). The Vapnik-Chervonenkis dimension, or VC-dimension of the set system $(S, \mathcal{R})$ is the maximum size of a shattered subset of $S$ ([VC]). The primary shatter function of a set system $(S, \mathcal{R})$ is defined as follows: $\pi_{\mathcal{R}}(m) = \max_{A \subset S, |A|=m} |\{R \cap A | R \in \mathcal{R}\}|$. A result by [VC] gives a bound on the shatter function of a set system with VC-dimension $\alpha$: $\pi_{\mathcal{R}}(m) = O(m^{\alpha})$. Intuitively the VC-dimension of a set system shows how complicated it is to describe.

Set systems with finite VC-dimension occur naturally in geometry and in learning theory. Lemma 1.a has been also shown by Welzl ([W]).

*Lemma 1.a:* The VC-dimension of the set system $([0,1]^2, \mathcal{F}_k([0,1]^2))$ is $2k+1$.

*1.b:* The VC-dimension of the set system $(S, \mathcal{F}_k(S))$ is equal to the minimum of the size of the maximum convex subset of $S$ and $2k+1$, and can be computed in $O(|S|^3)$ time.

*Proof:* *a:* Let us first show that no subset of $[0,1]^2$, with more than $2k+1$ points can be shattered. Take any $T$, $T \subset [0,1]^2$, $|T| = 2k+2$. Find the convex hull of $T$. If a point of $T$ is not in the convex hull, then, after we triangulate the convex hull this point is in the interior of a triangle. Any convex polygon that includes the corners of the triangle has to include the interior point as well; therefore $T$ is not shattered. If all the points of $T$ are on the convex hull, then we can partition them in two independent sets of $k+1$ points each. A convex polygon that includes exactly one of the two sets cannot have less than $k+1$ edges (Fig. 2.a); again $T$ is not shattered. Now let $T$ be a convex set with $2k+1$ points. Because the set is convex, a single halfspace can separate any subset $T'$ of consecutive points and $T \setminus T'$ (Fig. 2.b). It is then easy to see that $k$ halfspaces are sufficient to separate any subset of $T$, and so $T$ is shattered.

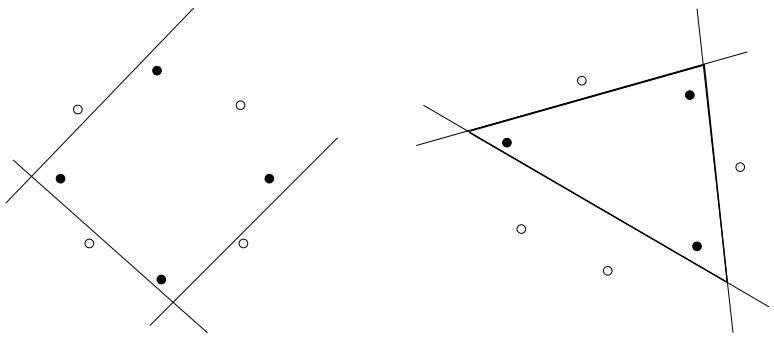

Figure 2: Three halfspaces cannot shatter 8 points but they can shatter 7.

*b:* From the first part we know that a maximal shattered subset has to be convex, and have cardinality at most $2k+1$. Edelsbrunner and Guibas ([EG]) give an algorithm that computes the maximum convex subset of a two dimensional point set $S$ in $O(|S|^3)$ time. □

The bound on the shatter function shows that there are at most $O(n^{2k+1})$ sets of points that we have to consider in order to find $\Delta_{max}(S, \chi, \mathcal{F}_k)$. The following lemma shows how to find them.

*Lemma 2:* For given $S \in [0,1]^2$ and $\chi$, There exists a polygon $A$ with at most $k$ edges such that $\Delta(A) = \Delta_{max}(S, \chi, \mathcal{F}_k(S))$ and each edge of $A$ passes through two points of $S$.

*Proof:* Take a maximal $k$-gon $A$, and choose an edge. Translate it parallel to itself until it touches a point. Then, rotate it around the attached point until a second point is encountered (Fig. 3). Repeating this operation for each edge we obtain $A'$. Each operation cannot not increase the number of edges of $A$. If a point $p$ on the boundary of $A'$, we assume it is in the interior of $A'$ if $\chi(p) > 0$, otherwise we assume it is in the exterior. This assumption increases the discrepancy, and can always be realized by a $k$-gon if no 3 points are collinear. So during this transformation the discrepancy of $A'$ can only increase, or $\Delta(A') \geq \Delta(A)$. Since $A$ maximizes the discrepancy, $\Delta(A')$ has to be maximum too. □

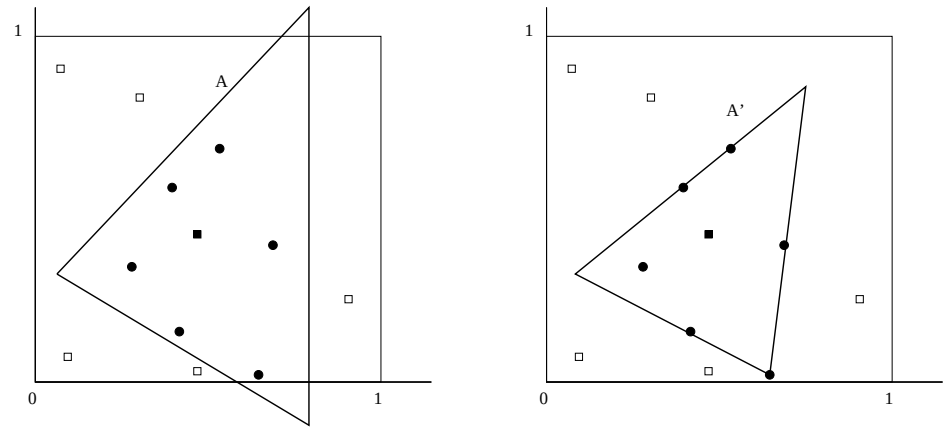


Figure 3: The discrepancy of $A'$ is equal to that of $A$.

*Lemma 3:* The bichromatic discrepancy $\Delta_{max}(S, \chi, \mathcal{F}_k(S))$ of a set $S$ ($S \subset [0,1]^2$, $|S| = n$) and a weight function $\chi$ ($\chi : S \rightarrow \Re$) for the set of convex $k$-gons can be computed in $O(n^{\alpha+1/2} \log n)$, where $\alpha =$ VC-dim $(S, \mathcal{F}_k(S))$.

*Proof:* Assume $S$ and $\chi$ are given. First we compute $\alpha =$ VC-dim $(S, \mathcal{F}_k(S))$ in $O(n^3)$ time. We consider all subsets of $S$ with at most $\alpha$ points. For each, we find its convex hull. Chazelle and Welzl ([CW]) show how to construct an algorithm that computes the weight of points inside a $k$-gon in $O(n^{1/2} \log n)$ time, provided a spanning path of low stabbing number ($O(n^{1/2})$) is known. Matousek ([Mat]) shows that we can construct such a spanning path in $O(n^{1+\delta})$ time ($\delta$ depends on the $O$ constant). We use this algorithm to find the number of

points of $S$ in the interior of the convex hull. We do two separate queries for positive and negative points. If $\alpha \leq k$ then obviously this subset can be in the interior of a $k$-gon, but if $\alpha > k$ we have to determine if there exists a $k$-gon that contains exactly these $\alpha$ points and the points in the interior of their convex hull. To do so, we find the $O(\alpha^2)$ halfspaces that the $O(\alpha)$ points in the convex hull define, and check, in $(O(\alpha^{2k} n^{1/2} \log n))$ time, if any of the $O(\alpha^{2k})$ $k$-gons thus formed contains exactly the same number of points with the convex hull of the $\alpha$ points. To increase the discrepancy, we assume that from the points that lie on the boundary of any $k$-gon, only the positive ones are in the interior and contribute to its discrepancy (Fig. 4).

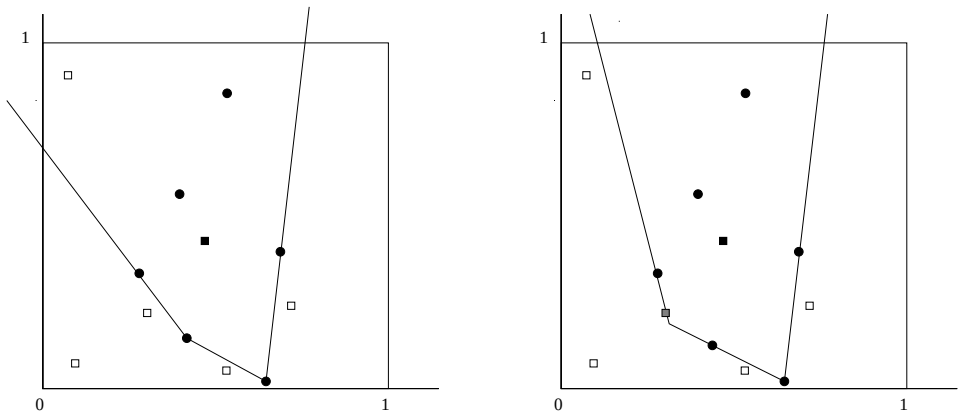


Figure 4: The subset on the left cannot be separated by 3 halfspaces that pass through its points. But the subset on the right has at least equal discrepancy.

Since $\alpha$ is the size of the largest convex subset of $S$, for every $k$-gon the points in its interior have a convex hull of at most $\alpha$ points. Lemma 2 shows that there exists an $F \in \mathcal{F}_k$ that maximizes the discrepancy and each edge passes through two points. These points must be in the convex hull of $F \cap S$, and therefore $F$ is considered by the algorithm. It follows that this algorithm computes $\Delta_{max}(S, \chi, \mathcal{F}_k(S))$ in $O(n^{\alpha+1/2} \log n)$ time. □

It is clear that this algorithm is not optimal when $\alpha = 2k + 1$. Indeed we can enumerate all $k$-gons that have each edge passing through a point in $O(n^{2k})$ time, and compute the discrepancy of each $k$-gon in linear time. Lemma 2 shows that this $O(n^{\alpha})$ algorithm gives the correct solution. In the next section we will show how to improve this simple algorithm.

# 3. The maximum bichromatic discrepancy of $k$-gons

First let us consider the case where our set of ranges is the set of the halfspaces ($\mathcal{F}_1$). A dynamic algorithm for this problem will be useful in the general case. The problem of

computing the numerical discrepancy of halfspaces was addressed by Dobkin and Eppstein ([DE]), and a dynamic algorithm for the numerical discrepancy is given. It is easy to show the following lemma if we follow their ideas and use the dynamic one-dimensional algorithm (given in [DGM]) that computes the bichromatic discrepancy of a point set in $O(\log n)$ time per insertion or deletion.

*Lemma 4:* We can insert and delete points in a set $S \subset [0,1]^2$ with an arbitrary coloring $\chi : S \rightarrow \{-1, +1\}$, and recompute the maximum halfspace bichromatic discrepancy in $O(n \log n)$ time, and using $O(n^2)$ space.

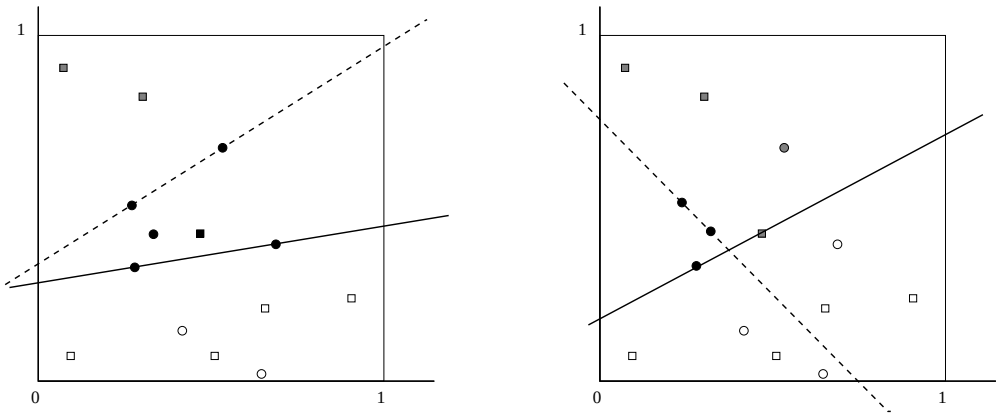


Figure 5: We use the dynamic algorithm for the halfspace discrepancy to find the new maximum wedge (circles are mapped to +1, and squares are mapped to -1, the filled points are the ones in the interior; the dotted line is in each case the maximum halfspace).

This dynamic algorithm can now be used to find the $k$-gon that maximizes the discrepancy. Let us fix $k-2$ halfspaces. Let $S'$ be the subset of $S$ in the intersection of those halfspaces. If we find the wedge that maximizes the discrepancy for $S'$ for all possible choices of $k-2$ halfspaces, we find the $k$-gon that maximizes the discrepancy.

For a given point $p \in S'$, we sort the points of $S' \setminus \{p\}$ counter-clockwise around $p$. We then find the set of points $S'_p = \{q | q \in S \text{ and } q = (q.x, q.y) \text{ and } q.y > p.y\}$. This is the set of the points of $S'$ that lie in the top halfspace that is defined by the horizontal line that passes from $p$. Using the dynamic algorithm, we find the halfspace that maximizes the discrepancy for $S'_p$. We then sweep the horizontal line counter-clockwise, appropriately inserting or deleting the points we encounter, and using the dynamic date structure to find the new maximizing halfspace. Thus, in $O(nn \log n) = O(n^2 \log n)$ time we find the wedge that passes from $p$ and maximizes the discrepancy (Fig. 5). We repeat the process for all points, and thus find the maximizing wedge. So we can find the $k$-gon that maximizes the discrepancy in $O(n^{2(k-2)} n^3 \log n) = O(n^{2k-1} \log n)$ time, and $O(n^2)$ space.

This algorithm, and lemma 3, give the following theorem:

*Theorem 1:* The bichromatic discrepancy $\Delta_{max}(S, \chi, \mathcal{F}_k(F))$ of a set $S$ ($S \subset [0,1]^2$, $|S| = n$) and a weight function $\chi$ ($\chi : S \rightarrow \Re$) for the set of convex $k$-gons can be computed in $O(n^{\min(\alpha+1/2, 2k-1)} \log n)$, where $\alpha =$ VC-dim $(S, \mathcal{F}_k(S))$.

# 4. The maximum bichromatic discrepancy of stripes

In this section we examine a problem that is similar, but allows a very different approach. Let us give some notation first. A stripe can be defined as the intersection of two halfspaces with parallel supporting lines. Let's assume that $l_1$ and $l_2$ are parallel lines. We name the stripe they form $s(l_1, l_2)$. Let $\mathcal{F}_s$ be the set of planar stripes. The problem is to compute $\Delta_{max}(S, \chi, \mathcal{F}_s(S))$ for given $S, \chi$. The following lemma, similar to lemma 2, shows that we have to consider only a finite number of stripes.

*Lemma 5:* There exists a stripe $s(l_1, l_2)$ that maximizes $\Delta$ such that $l_1$ and $l_2$ are parallel to one of the $n(n-1)/2$ lines defined by all pairs of points of $S$.

A simple algorithm to compute $\Delta_{max}(S, \chi, \mathcal{F}_s(S))$ is then to find, for each of the $O(n^2)$ possible pairs of points, the stripe which is parallel to the line and maximizes $\Delta$. Our algorithm does exactly that.

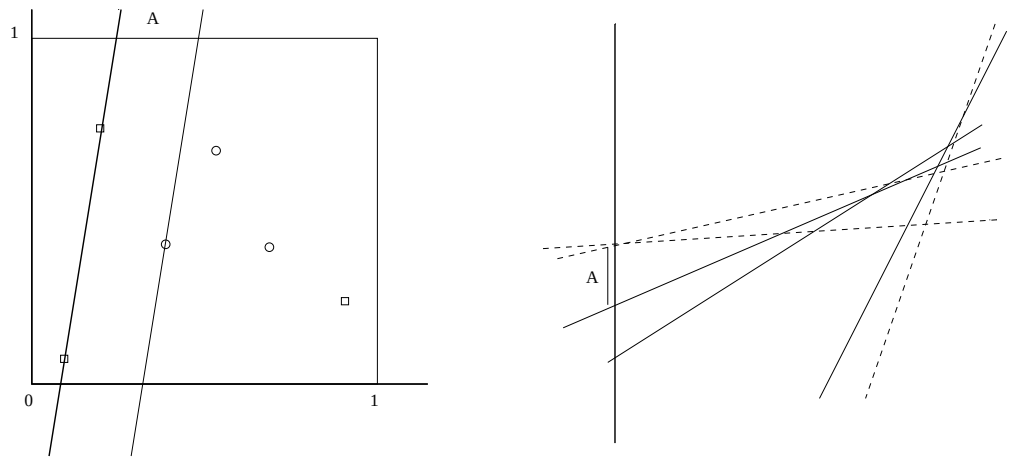


Figure 6: A stripe and its dual.

To do so we use a dual space transformation. The basic idea is to create a mapping $\mu$ between points in the primal space and non-vertical lines in the dual space. A point $(a, b)$ in the primal space is mapped to the line $\mu(a, b) = (y = a\,x + b)$ in the dual space. The dual of a line $y = a\,x + b$ is then the point $\mu(y = a\,x + b) = (-a, b)$ in dual space. An important property of this transformation is that if the point $(a_1, b_1)$ is above the line $y = a_2\,x + b_2$ in primal space (that is $(b_1 > a_2\,a_1 + b_2)$, then the line it is mapped to $(y = a_1\,x + b_1)$ is above the point $(-a_2, b_2)$ (because $b2 < -a_1\,a_2 + b_1$).

In the dual space, the $n$ points of $S$ are mapped to $n$ lines. The resulting arrangement has $O(n^2)$ vertices and regions, and it can be constructed in $O(n^2)$ time ([DM]). The following lemma shows that a non-vertical stripe in the primal space maps to a vertical line segment in the dual space.

*Lemma 6:* Let $s(l_1, l_2)$ be a stripe in primal space, such that $l_1$ is not vertical. Then the line segment $\mu(l_1)\mu(l_2)$ is vertical, and, for every $p \in S$, $p \in s(l_1, l_2)$ if and only if $\mu(l_1)\mu(l_2)$ intersects $\mu(p)$.

*Proof:* Since $l_1$ and $l_2$ are parallel and not vertical, there exist $a, b, c$ such that $l_1$ can be expressed as $y = a\,x + b$ and $l_2$ can be expressed as $y = a\,x + c$ (WLOG also assume that $b > c$). Then $\mu(l_1) = (a, b)$ and $\mu(l_2) = (a, c)$, so $\mu(l_1)\mu(l_2)$ is vertical. Take a point $p_i = (x_i, y_i) \in S \cap s(l_1, l_2)$. Since $l_1$ is above $l_2$, $p_i$ must be below $l_1$ (that is $y_i < a\,x_i + b$) and above $l_2$ ($y_i > a\,x_i + c$). Then in the dual space, the line $\mu(p)$ must pass below the point $\mu(l_1)$ and above the point $\mu(l_2)$, it must therefore cross the line segment $\mu(l_1)\mu(l_2)$. Conversely, a line that crosses $\mu(l_1)\mu(l_2)$ in the dual space corresponds to a point that lies below $l_1$ and above $l_2$ in the primal space, so it must be in the interior of $s(l_1, l_2)$. □

Let us take a vertical crosssection of the dual arrangement at some $x$-coordinate $a$. A line segment on this crosssection defines a stripe parallel to the line $y = (-a)\,x$. The vertical crosssection intersects the dual lines, and we get an one-dimensional set $S'$ of $m = n(n-1)/2$ points. From lemma 5, the line segment that maximizes the discrepancy for this set $S'$ and the original $\chi$ defines the stripe that maximizes the discrepancy among all stripes parallel to $y = (-a)\,x$. This one dimensional problem has been studied before, and [DGM] give an $O(\log m)$ time and linear space dynamic algorithm.

Lemma 2 defines the only $n(n-1)/2$ $x$-coordinates that we have to consider. They are the lines defined by pairs of points of $S$, and in the dual space they map to the points of the arrangement. To solve the $O(n^2)$ problems efficiently, we sort the $m$ dual points on their $x$-coordinate, and we do a horizontal sweep of the arrangement. In the computation of the bichromatic discrepancy at each snapshot, what matters is the relative position of the points, and not the actual $y$-coordinates. So, when we find $S'$ at the leftmost arrangement point, we replace the $y$-coordinates with ranks.

*Theorem 2:* We can find the maximum stripe bichromatic discrepancy of a set $S \subset [0, 1]^2$

with an arbitrary coloring $\chi : S \to \Re$, in $O(n^2 \log n)$ time, and in $O(n^2)$ space. [1]

# 5. Conclusions and extensions

A very interesting open problem is giving practical algorithms for higher dimensions. Some of the algorithms we present extend to 3 and higher dimensions, but the running times become much worse.

We consider the special case of tetrahedra, but all the arguments extend to convex $k$-hedra (for constant $k$). It is not difficult to show that there exists a tetrahedron that maximizes the discrepancy, and has 3 points on each face, and such a tetrahedron can be computed in $O(n^{11} \log n)$ time. We give a sketch of the algorithm here. Extending Lemma 4 to three dimensions, we can dynamically compute the halfspace that maximizes the discrepancy in $O(n^2 \log n)$ time. To compute the maximum intersection of two halfspaces we sweep around every pair of points, in $O(n^2\ n\ n^2 \log n)$. This has to be done for every set of points that is in the intersection of two halfspaces.

It is interesting to observe that the VC dimension of tetrahedra is at most 16 and at least 14 ($14 \leq$ VC-dim $([0,1]^3, \mathcal{F}_4([0,1]^3)) \leq 16$). Indeed a convex set of 17 points has always an independent set of vertices of size 5, because the corresponding planar graph can be four-colored ([W]). However 4 hyperplanes cannot include exactly the remaining 12 points because each one can separate only one of the 5 points of the independent set. Therefore a convex set of 17 points cannot be shattered. In addition, lemma 1 also shows that a set of points in non-convex position cannot be shattered either, and so the VC-dimension cannot exceed 16. On the other hand, the 12 vertices of a regular eikosahedron can be shattered because there exist 4 independent triangular faces, and a halfspace can separate any subset of the 3 points of each such face. Two more points can be positioned in so that the choice on the independent set of faces determines if they are on the inside or the outside of the tetrahedron (Fig. 7). The VC-dimension of $(S, \mathcal{F}_4(S))$ with $S \subset [0,1]^3$ cannot be larger than the size of the maximum convex subset $S_c \subset S$ either.

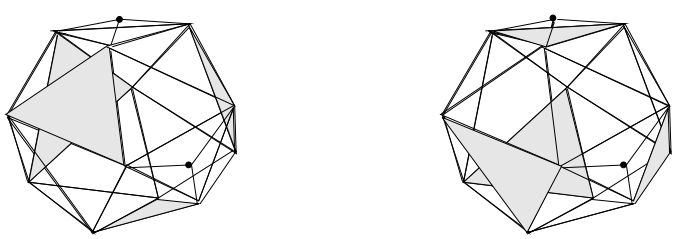

Figure 7: VC-dim$([0,1]^3, \mathcal{F}_4([0,1]^3)) \geq 14$.

[1] We note here that [dB] gives a more complicated algorithm to compute the numerical discrepancy of stripes in $O(n^2 2^{a(n)} \log n)$.

Other important open problems include showing if these algorithms are optimal, and finding good approximations in reasonable time bounds.

The algorithms we present for computing the maximum discrepancy have direct applications in learning theory and machine learning, and their running time allows them to be useful with the large data sets common in many real-life problems. Finally, our techniques can also be applied to the computation of the numerical discrepancy.

# References


[AL] D. Angluin and P. Laird, *Learning from noisy examples,* Machine Learning 2 (1988), 343-370.

[BC] J. Beck and W.W.L. Chen, *Irregularities of distribution,* Cambridge University Press (1987).

[BN] W. Buntine and T. Niblett, *A further comparison of splitting rules for decision-tree induction,* Machine Learning 8 (1992), 75-82.

[C93a] B. Chazelle, *Geometric Discrepancy Revisited,* 34th IEEE Symp. Foundat. Computer Science (1993).

[C93b] B. Chazelle, *Cutting hyperplanes for divide-and-conquer,* Disc. Comput. Geom. 9, 145-158 (1993).

[CW] B. Chazelle and E. Welzl, *Quasi-optimal Range Searching in Spaces of Finite VC-Dimension,* Disc. Comput. Geom. 4, 467-489 (1989).

[dB] M. deBerg, *Computing Half-Plane and Strip Discrepancy of Planar Point Sets,* to appear (1994).

[DE] D. Dobkin and D. Eppstein, *Computing the Discrepancy,* 9th Ann. Symp. on Comput. Geom. (1993), 47-52.

[DM] D. Dobkin, and D. Mitchell, *Random-Edge Discrepancy of Supersampling Patterns,* Graphics Interface (1993).

[DGM] D. Dobkin, D. Gunopulos and W. Maass, *Computing the Bichromatic Discrepancy, with applications in Computer Graphics and Machine Learning,* JCSS, to appear.

[EG] H. Edelsbrunner and L. J. Guibas,*Topologically Sweeping an Arrangement,* JCSS 38, 165-194 (1989).

[Ha] D. Haussler, *Decision theoretic generations of the PAC-model for neural nets and other applications,* Inf. and Comp. 100 (1992), 78-150.

[HW] D. Haussler and E. Weltz, *$\epsilon$-nets and simplex range queries,* Disc. Comput. Geom. 2 (1987), 127-151.

[Ho] R.C. Holte, *Very simple classification rules perform well on most commonly used datasets,* Machine Learning 11 (1993), 63-91.

[Ma] W. Maass, *Efficient Agnostic PAC-Learning with Simple Hypotheses,* 7th Ann. ACM Conf. on Computat. Learning Theory (1994), 67-75.

[Mat] J. Matusek, *Efficient Partition Trees,* 7th Ann. ACM Symp. on Comput. Geom. 1-9 (1991).

[MWW] J. Matousek, E. Welzl and L. Wernish, *Discrepancy and $\epsilon$-approximations for bounded VC-dimension,* 25th ACM Symp. on Theory of Computing (1993), 424-430.

[Mi] J. Mingers, *An empirical comparison of pruning methods for decision tree induction,* Machine Learning 4 (1989), 227-243.

[VC] V.N. Vapnik and A.Ya. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities,* Theory Probab. Applic. 16 (1971), 264-280.

[WGT] S.M. Weiss, R. Galen and P.V. Tadepalli, *Maximizing the predictive value of production rules,* Art. Int. 45 (1990), 47-71.

[WK90] S.M. Weiss and I. Kapouleas, *An empirical comparison of pattern recognition, neural nets, and machine learning classification methods,* 11th Int. Joint Conf. on Art. Int. (1990), Morgan Kauffmann, 781-787.

[WK91] S.M. Weiss and C.A. Kulikowski, *Computer Systems that Learn,* Morgan Kauffmann (1991).

[W] E. Welzl, personal communication.