Dimensionality Reductions that Preserve Volumes and Distance to Affine Spaces, and their Algorithmic Applications

Avner Magen

NEC Research Institute, Princeton
avner@research.nj.nec.com.

Abstract. Let X be a subset of n points of the Euclidean space, and let $0 < \varepsilon < 1$. A classical result of Johnson and Lindenstrauss [JL84] states that there is a projection of X onto a subspace of dimension $O(\varepsilon^{-2} \log n)$, with distortion $\leq 1 + \varepsilon$. Here we show a natural extension of the above result, to a stronger preservation of the geometry of finite spaces. By a k-fold increase of the number of dimensions used compared to [JL84], a good preservation of volumes and of distances between points and affine spaces is achieved. Specifically, we show it is possible to embed a subset of size n of the Euclidean space into a $O(\varepsilon^{-2} k \log n)$- dimensional Euclidean space, so that no set of size $s \leq k$ changes its volume by more than $(1+\varepsilon)^{s-1}$. Moreover, distances of points from affine hulls of sets of at most $k-1$ points in the space do not change by more than a factor of $1+\varepsilon$. A consequence of the above with $k = 3$ is that angles can be preserved using asymptotically the same number of dimensions as the one used in [JL84]. Our method can be applied to many problems with high-dimensional nature such as *Projective Clustering* and *Approximated Nearest Affine Neighbor Search*. In particular, it shows a first poly-logarithmic query time approximation algorithm to the latter. We also show a structural application that for volume respecting embedding in the sense introduced by Feige [Fei00], the host space need not generally be of dimensionality greater than polylogarithmic in the size of the graph.

1 Introduction

The dimension of a normed space that accommodates a finite set of points plays a critical role in the way this set is analyzed. The running time of most geometric algorithms is at least linear in the dimensionality of the space, and in many cases exponential in it. To represent the metric of n points in the Euclidean space, one clearly needs no more than $n-1$ dimensions. By relaxing the notion of isometry to near-isometry, the underlying structure of these points is well represented in a space with a much smaller number of dimensions; In their seminal paper [JL84], Johnson and Lindenstrauss show that far more efficient representations capture almost precisely the metric nature of such sets. They present a simple and elegant principle that allows one to embed such an n-point set into a t-dimensional Euclidean space, with t merely $O(\varepsilon^{-2} \log n)$, while preserving the pairwise distances to within a relative error of ε. One simply needs to project the original space onto a random t-dimensional subspace (and scale accordingly) to obtain a low-distortion embedding with high probability. Their argument therefore supplies a probabilistic algorithm for producing such an embedding.

Modifications that relate to the exact way the randomization is applied, and further improvements in the parameters were later proposed. For example it was shown that by using a projection onto t independent random unit vectors, a similar result can be achieved. In [Ach01], Achlioptas shows an even simpler probability space for the desired embedding, by projecting the space onto random vectors in $\{-1,1\}^N$. It is interesting to note that these simplifications do not require higher dimensionality to satisfy the same quality of embeddings. In [EIO02], a derandomization of the probabilistic projection is given, leading to an efficient deterministic algorithm for finding such low-dimensional embeddings. Among the other simpler proofs to the result of Johnson and Lindenstrauss (which we sometimes call *JL-lemma*) are [FM88,IM98,LLR95,DG99,AV99].

For metric spaces (X, d_X) and (Y, d_Y), and an embedding $f : X \to Y$, we define the distortion of f by $\sup_{x,y \in X} \frac{d_Y(f(x),f(y))}{d_X(x,y)} \cdot \sup_{x,y\in X} \frac{d_X(x,y)}{d_Y(f(x),f(y))}$. By this definition a good embedding is one that preserves the pairwise distances. However, a set X of points in the Euclidean space has many more characteristics in addition to the metric they represent, such as the center of gravity of a set of points and its average distance to them, the angles defined by triplets of points, volumes of sets, and distances between points to lines, planes and higher dimensional affine spaces that are spanned by subsets of X. Regarding (some of) these characteristics as part of the structure of X, we redefine the quality of an embedding to be one that preserves both the volumes of (certain) subsets of X and the distances of points from affine hulls of subsets of X.

We define the volume of a set of k points in the Euclidean space as the $(k-1)$-dimensional volume (Lebesgue measure) of its convex-hull. For $k = 2$ this is just the distance between the points. For $k = 3$, this is the area of the triangle with vertices that are the three points of the set, etc. Throughout this paper we denote the volume of a set S in the Euclidean space by $\mathrm{Vol}(S)$. When considering a general metric space (not necessarily Euclidean), it is not a-priori clear whether there is a reasonable way to define a volume. In [Fei00] Feige defined a notion of volumes for general metric spaces, and measured the quality of an embedding from general metric spaces into Euclidean spaces (he calls such embeddings *volume respecting embeddings*. The volume preservation there applied to two different definition of volumes, the one in general metric spaces, and the one in Euclidean space. This line of work led to important algorithmic applications, most notably a polylogarithmic approximation algorithm for the bandwidth problem [Fei00], and an approximation algorithm to a certain VLSI layout problem [Vem98]. Our attention focuses on the case where both the original and the image space are Euclidean, and consequently the volume preservation notion is a well defined one and need not use the more involved definition of [Fei00]. Accordingly, our result should not be confused with results in the aforementioned framework, most notably Rao's result [Rao99], that deals with Euclidean metric, but with respect to its metric structure alone.

Consider an embedding $f : \mathbb{R}^N \to \mathbb{R}^t$ that does not expand distances in $X \subset \mathbb{R}^N$. We say f distorts the volume of a set $S \subseteq X$ of size k by $\left(\frac{\mathrm{Vol}(S)}{\mathrm{Vol}(f(S))}\right)^{\frac{1}{k-1}}$. The exponent in this expression should be thought of as a natural normalization measure (and was introduced in [Fei00]).

Our Result: Let $\varepsilon \leq \frac{1}{4}$, X be an n-point subset of $\mathbb{R}^N$, and let $t = O(\varepsilon^{-2} k \log n)$. We show that there is a mapping of $\mathbb{R}^N$ into $\mathbb{R}^t$ that (i) does not distort the volume of subsets of X of size at most k to by more than a factor of $1 + \varepsilon$. (ii) preserves the distance of points from affine hulls of subsets of X of size at most $k - 1$ to within a relative error of ε.

To see how our result is achieved, we take a closer look into JL-lemma. The JL-lemma is based on the following lemma that can be found (in slightly different formulations) in [Ach01,DG99,IM98].

Lemma 1. *Let $\varepsilon \leq \frac{1}{3}$, $v \in \mathbb{R}^N$, and let f be a random projection onto t dimensions, multiplied by $\sqrt{\frac{N}{t(1+\varepsilon)}}$. Then*

$$\mathrm{Prob}\left(\frac{\|v\|}{1+\varepsilon} \leq \|f(v)\| \leq \|v\|\right) \geq 1 - \exp(\frac{2}{15} t \varepsilon^2)$$

For a low distortion embedding, $\binom{n}{2}$ vectors (the unsigned pairwise differences between the points) should maintain their norms approximately, and so t is chosen so that the above probability is smaller than $1/\binom{n}{2}$ resulting in $O(\varepsilon^{-2} \log n)$ needed dimensions.

In this paper we show that in order for a linear embedding to preserve volumes and affine-distances of sets of size at most k to within a relative error of ε, it is sufficient to preserve the norms of a certain set of $\exp(O(k \log n))$ vectors, to within a relative error of $\varepsilon/3$. The result is then achieved by taking t to be $O(\varepsilon^{-2} k \log n)$ which guarantees a positive probability for the preservation of the norms of that many vectors.

Applications In [IM98], Indyk and Motwani describe the way projections can be used for designing efficient algorithms for the *Approximate Nearest Neighbor* problem. The generalization of this problem from a set of points to a set of k-dimensional affine spaces is the *Approximate Nearest affine neighbor*. For $k = 1$, i.e., this is the problem of finding the closest line to a query point. This is a natural proximity problem, that has appeared few times in the literature. Using Meiser's result for point location in arrangements of hyperplanes together with our result yields a randomized poly-logarithmic query-time approximation to the *Approximate Nearest affine neighbor* problem, which is to the best of our knowledge the first. In Section 5.1 we show how exactly this can be achieved. Our result can be applied to another classical problem in computational geometry that stems from data mining. Consider a data set in $\mathbb{R}^N$, where the "true dimensionality" is anticipated to be much smaller than N. In other words, it is assumed to be possible to cover (up to proximity) the set by a small number of k-dimensional affine spaces. The problem is known to be NP-hard even for dimension

2. Currently, there are approximation algorithms for the cases of two and three dimensions. For the higher dimensions, our method provides a way to reduce the problem to $\varepsilon^{-2}k\log n$ dimensions.

There are also straight forward applications of our result to other problems which extend metrical questions to ones that consider volumes. In particular, the problem of finding the diameter of a set of n points, can be extended to finding the biggest volume subset of size k. It is immediate by our result that this can be approximated when dimensionality is reduced to $O(k\log n)$.

Can Feige's volume-respecting embeddings benefit from our result? We briefly describe the general framework in those embeddings. Consider an embedding of a graph on n vertices into the Euclidean space that does not increase distances. Such an embedding is good if the (Euclidean) volumes of sets of size at most k are big. Typically the value of k is $O(\log n)$. Our result shows that by combining a volume-respecting embeddings with a random projection onto a low dimension, the distortion is asymptotically the same. Specifically, for $k = O(\log n)$ as often is the case, we get that the restriction to $O(\log^2 n)$-dimensional embeddings entails only an extra constant factor to volume-distortion of the embedding.

2 Preliminaries and Notation

The norm $\|\cdot\|$ always stands for the Euclidean norm. We say that an embedding $\phi : X \to \mathbb{R}^N$ is a *contraction* if for every $x, y \in X$, $\|\phi(x) - \phi(y)\| \leq \|x - y\|$. Whenever the dimensionality of an Euclidean space is immaterial, we call it $\mathbb{R}^N$ without explicitly defining N. We will sometime refer to an affine space as a *flat*. An affine subspace of $\mathbb{R}^N$ that is spanned by points of $X \subseteq \mathbb{R}^N$ is called X-flat (analogously, we define X-lines, X-planes, X-k-dimensional flats, etc.).

For a set $S \subseteq \mathbb{R}^N$ of size ν, we denote by $\mathcal{L}(S)$ the affine-hull of S, that is $\mathcal{L}(S) = \{\sum_{i=1}^{\nu} \lambda_i a_i | \sum_i \lambda_i = 1\}$. For a set $S \subseteq \mathbb{R}^N$ and $x \in \mathbb{R}^N$ we define $P(x, S)$ to be the projection of x onto $\mathcal{L}(S)$. The *affine distance* of x to S, $\mathrm{ad}(x, S)$, is defined to be the distance of x to the affine-hull of S, or equivalently $\|x - P(x, S)\|$. The affine distance will occasionally referred to as *height*. Let $r_1, r_2 \ldots, r_{\nu-1}$ be an arbitrary set of orthonormal vectors in $\mathcal{L}(S)$[1]. We now define the *corner* points of the pair (x, S). The i-th corner point $c_i(x, S)$ is defined as $P(x, S) + \mathrm{ad}(x, S) \cdot r_i$, for $1 \leq i \leq \nu - 1$. When $s = 2$ we let $c(x, S)$ denote $c_1(x, S)$. See figure 3.

3 Preserving Distances to Lines and Preserving Areas of Triangles

Consider the problem of finding a low-dimensional Euclidean embedding of a finite subset X of the Euclidean space, such that pairwise distances, areas of triangles, and distance of points from X-lines do not change by much. Note that this is a special case of the the general problem we consider (here $k = 3$), as volumes of triplets of points are simply areas. This case is easier to analyze and moreover, the analysis of the general case uses some of the structure of the two dimensional case. The current case also gives a preservation result for angles as we later note.

The natural thing to try in order to reduce dimensionality and preserve geometrical features is simply to apply JL-lemma: after all, in an isometry not merely distances are preserved, but also volumes, affine distances and angles. One might expect that when f is nearly an isometry (i.e. f has small distortion) it will follow that volumes and affine distances are being quite reasonably preserved. We show that in general, this is very far from the truth: Consider a triangle with a very small angle. A low-distortion embedding can be applied to it, so that it is changed to a triangle with dramatically different angles, area and heights (see Figure 1).

We next show a certain class of triangles for which small distance-distortion *does* imply small heights-distortion.

Lemma 2. *Let A, B, C be the vertices of a right angle isosceles triangle, where the right angle is at A, and let Φ be a contracting embedding of its vertices to a Euclidean space, such that the edges do not contract by*

[1] We exclude the case where S is affinely dependent: Since we consider only linear mappings, the image of affinely dependent set will also be affinely dependent, and so degenerated case will remain degenerated in the image. Also notice that $\mathrm{ad}(x, S) = \mathrm{ad}(x, S')$ where $S' \subseteq S$ is an affinely independent set for which $\mathcal{L}(S') = \mathcal{L}(S)$, and that if $\mathrm{ad}(x, S) = 0$ then $\mathrm{ad}(f(x), f(S)) = 0$.

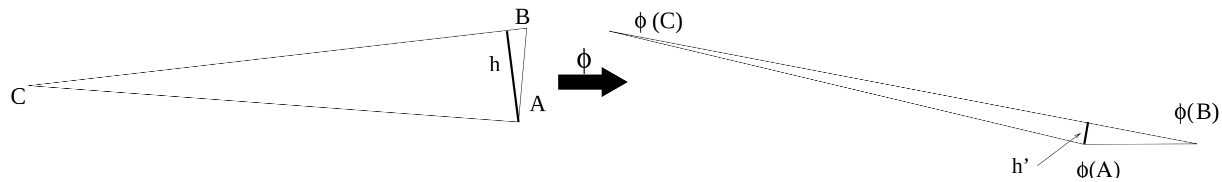

Fig. 1. Areas and affine distances can dramatically change and practically vanish, even under a low-distortion embedding. Here, the distortion of Φ is very small, but $h' \ll h$

more than $1+\varepsilon$, *where* $\varepsilon \le \frac{1}{6}$. *Let* h *be the length of* $[AC]$ $(h = \mathrm{ad}(C, \{A, B\}))$, $\boldsymbol{b}$ *be the vector* $\Phi(B) - \Phi(A)$, *and* $\boldsymbol{c}$ *be the vector* $\Phi(C) - \Phi(A)$. *Then*

1. $|\langle \boldsymbol{b}, \boldsymbol{c}\rangle| \le 2\varepsilon \cdot h^2$
2. $h/(1+2\varepsilon) \le \mathrm{ad}(\Phi(C), \{\Phi(A), \Phi(B)\}) \le h$.

Remark 1. The first assertion of the lemma says that the images of the perpendicular edges of the triangle are almost orthogonal too. This fact is used later in the analysis of the general case.

Proof. Let $a = ||\boldsymbol{b} - \boldsymbol{c}||$, and let θ be the angle between $\boldsymbol{b}$ and $\boldsymbol{c}$. Now, $|\langle \boldsymbol{b}, \boldsymbol{c}\rangle| = |||\boldsymbol{b}||^2 + ||\boldsymbol{c}||^2 - a^2|/2$. A simple analysis shows that this quantity is maximized (while satisfying the conditions on Φ) when $||\boldsymbol{b}|| = ||\boldsymbol{c}|| = h/(1+\varepsilon)$ and $a = \sqrt{2}h$. Hence $|\langle \boldsymbol{b}, \boldsymbol{c}\rangle| \le \frac{1}{2}h^2 \cdot (2 - 2/(1+\varepsilon)^2) \le 2\varepsilon \cdot h^2$. As can be easily verified, these values of $a, ||\boldsymbol{b}||, ||\boldsymbol{c}||$ also maximize $|\cos\theta| = |||\boldsymbol{b}||^2 + c^2 - a^2|/(2||\boldsymbol{b}||||\boldsymbol{c}||)$. Hence $|\cos\theta| \le |(2 - 2(1+\varepsilon)^2)/2| = 2\varepsilon + \varepsilon^2$, and accordingly $\sin\theta = \sqrt{1 - \cos^2\theta} \ge \sqrt{1 - 4\varepsilon^2 - 4\varepsilon^3 - \varepsilon^4} > \frac{1+\varepsilon}{1+2\varepsilon}$ for $\varepsilon \le \frac{1}{6}$. Finally,

$$h/(1+2\varepsilon) = \frac{h}{1+\varepsilon} \cdot \frac{1+\varepsilon}{1+2\varepsilon} \le \mathrm{ad}(\Phi(C), \{\Phi(A), \Phi(B)\}) = c \sin\theta \le h.$$

In order to conclude that under a low-distortion embedding of the set X, areas and affine distances do not change by much, one would like to eliminate bad cases such as those in Figure 1. Think of the following physical model: Edges are rubber rods that can slightly contract due to a shock. Figure 1 demonstrates that this shock may very well change the areas and heights of triangles significantly. The remedy we propose is to supplement this rubber triangle with some additional rods to keep it stable. If the contraction of these rods is also limited, then placing them in appropriate locations, will eliminate cases such as the mapping Φ in Figure 1. This mental experience translates to an additional set of vectors whose norms must be approximately preserved.

Our strategy is therefore choosing “rods” such that nice triangles as in Lemma 2 emerge. This, together with the fact the embeddings we consider are linear, enables us to bound the changes of the heights. Area preservation then immediately follows.

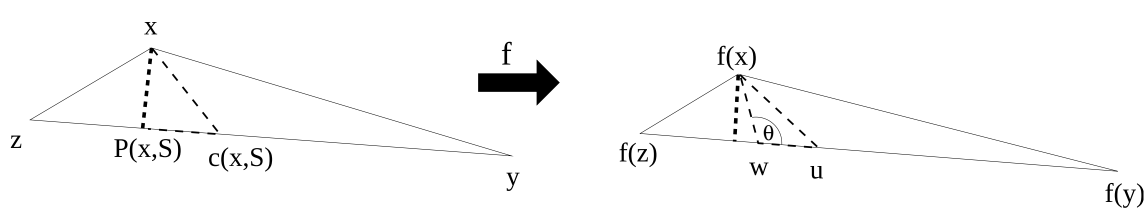

Fig. 2. The triangle $f(x), w, u$ is the f-image of a right angle isosceles triangle. Height-contraction is estimated via these triangle (dashed lines)

Theorem 1. *Let* $\varepsilon \le \frac{1}{3}$ *and let* n, t *be integers for which* $t \ge 60\varepsilon^{-2}\log n$. *Then for any* n*-point subset* X *of the Euclidean space* $\mathbb{R}^N$, *there is a linear contracting embedding* $f : X \to \mathbb{R}^t$, *under which the areas of* *triangles in X are preserved to within a factor of $(1+\varepsilon)^2$, the distances of points from X-lines are preserved to within a factor of $1+\varepsilon$, and angles (of triplets of points from X) are preserved to within a (double-sided) factor of $1+\frac{8}{\pi}\cdot\sqrt{\varepsilon}$.*

Proof. For every pair $S=\{y,z\}$ of elements of X, and every element $x\in X-S$, we consider the right angle isosceles triangle $\{x,P(x,S),c(x,S)\}$. Let V be the collection of the unsigned vectors corresponding to these triangles ($x-P(x,S), c(x,S)-P(x,S)$ and $x-c(x,S)$) over all choices of S and x, together with all pairwise differences between the points of X. Let f be a random projection onto t dimensions, multiplied by $\sqrt{\frac{N}{t(1+\varepsilon/2)}}$. By Lemma 1, the probability that f does not expand norms of vectors in V, and that it does not contract them by more than $1+\varepsilon/2$ is at least $1-|V|\exp(-\frac{2}{15}t(\frac{\varepsilon}{2})^2)$. A little closer look shows that V contains merely $3\binom{n}{2}$ different directions : the directions of X-lines, and their rotations by $\pi/4$ and by $\pi/2$. Since f is linear, vectors in the same direction are preserved simultaneously, and therefore to establish the existence of f as above we need

$$1-3\binom{n}{2}\exp\left(-\frac{2}{15}t(\varepsilon/2)^2\right)>0$$

which is satisfied when $t\geq 60\varepsilon^{-2}\log n$. Next we show that f satisfies the preservation statements of the theorem. Consider three different points $x,y,z\in X$. Let $w=f(P(x,\{y,z\})$ and $u=f(c(x,\{y,z\}))$. Now let θ be the angle between $f(x)-w$ and $f(y)-w$ (see Figure 2). Since $z,P(x,\{y,z\})$ are collinear and f is linear, $f(z),w,u,f(y)$ are also collinear. We now apply the second assertion of Lemma 2 with $P(x,\{y,z\}),c(x,\{y,z\})$ and x for A,B and C and $\varepsilon/2$ the error parameter. It follows that

$$\mathrm{ad}(x,\{y,z\})/(1+\varepsilon)\leq \mathrm{ad}(f(x),\{f(z),f(y)\})\leq \mathrm{ad}(x,\{y,z\}).$$

For the area estimates, we get

$$1/(1+\varepsilon)^2\leq\frac{\mathrm{Vol}(f(S))}{\mathrm{Vol}(S)}=\frac{\|f(y)-f(z)\|}{\|y-z\|}\cdot\frac{\mathrm{ad}(f(x),\{f(y),f(z)\})}{\mathrm{ad}(x,\{y,z\})}\leq 1.$$

Finally, let $\alpha=\measuredangle(xyz)$ and $\alpha'=\measuredangle(f(x)f(y)f(z))$. We have that

$$\frac{1}{1+\varepsilon}\leq\frac{\sin\alpha'}{\sin\alpha}=\frac{\|y-x\|}{\|f(y)-f(x)\|}\cdot\frac{\mathrm{ad}(f(x),\{f(y),f(z)\})}{\mathrm{ad}(x,\{y,z\})}\leq 1+\varepsilon. \tag{1}$$

We now turn to analyze the relative change of the angles themselves. For the case $\alpha,\alpha'\leq\pi/2$, we use the following fact that holds for any $0<\beta,\gamma\leq\pi/2$: if $\sin\beta/\sin\gamma\leq 1+\varepsilon$ then $\beta/\gamma\leq 1+\sqrt{\varepsilon}$. This fact together with inequality 1 implies that $1/(1+\sqrt{\varepsilon})\leq\alpha'/\alpha\leq 1+\sqrt{\varepsilon}$. The case where $\alpha,\alpha'\geq\pi/2$ can be easily reduced to the previous case, by replacing α,α' by $\pi-\alpha,\pi-\alpha'$. Last, assume that $\alpha\geq\pi/2$ and $\alpha'\leq\pi/2$. Using the fact that a linear embedding is an isometry times a scalar when restricted to a line, and in particular that the order of points along a line does not change under a linear embedding, we get $\|f(x)-w\|\geq\|f(x)-f(y)\|$. Now,

$$\|f(x)-w\|\leq\|x-P(x,\{y,z\})\|=\|x-y\|\cdot\sin\alpha\leq\frac{\|f(x)-f(y)\|}{1+\varepsilon/2}\cdot\sin\alpha,$$

and therefore $\sin\alpha\geq 1/(1+\frac{\varepsilon}{2})$ which means $\alpha\leq\pi/2+\sqrt{\varepsilon}$. Analogously, we can show that $\alpha'\geq\pi/2-\sqrt{\varepsilon}$. Therefore $\alpha-\alpha'\leq 2\sqrt{\varepsilon}$. When the angle changes from acute to obtuse we similarly obtain that $\alpha'-\alpha\leq 2\sqrt{\varepsilon}$. Summing it all up, we get

$$1/(1+\varepsilon')\leq\alpha'/\alpha\leq 1+\varepsilon',$$

where $\varepsilon'=\frac{8}{\pi}\cdot\sqrt{\varepsilon}$ always holds.

4 The General Case

Analogously to the $k=3$ case, we look for a set of vectors with the following properties: (i) it is not too big and (ii) if a linear embedding does not change the norms of the vectors by much, then it also does not change affine distances and volumes by much. We first restrict our attention to one affine distance, $\mathrm{ad}(x,S)$

with $|S| = s < k$. We construct a set of vectors in a way which is determined by the relation between ε and s. When s is small with respect to $1/\varepsilon$ we extend the previous construction (for the case $k = 3$) in the natural way. The right angle isosceles triangle is substituted by a simplex spanned by $[x, P(x, S)]$ and by orthogonal vectors in $\mathcal{L}(S)$ of size $\mathrm{ad}(x, S)$ (call this a *nice simplex*). When s is bigger than $1/\varepsilon$ we use, in addition to this set of vertices, a dense enough set of points. The analysis is mostly algebraic one. It relates to the geometry of interest in the following simple way: Let V_{s-1} be a simplex on s points, V_{s-2} be the facet of V_{s-1} opposite one of the points, and H be the distance from that point to the facet, then $\mathrm{Vol}(V_{s-1}) = H \cdot \mathrm{Vol}(V_{s-2})/(s-1)$. The simplices we take are the images of the nice simplices under the linear low distortion embedding. The actual estimate $\mathrm{Vol}(V_{s-1})/\mathrm{Vol}(V_{s-2})$ is then achieved by means of analysis of determinants of matrices with constraints resulting from the low distortion embedding on the set of auxiliary points we added.

Similarly to the case $k = 3$, we make use of the linearity of our embeddings to claim that it is enough to bound the contraction of heights in nice simplices to get the actual bound for all the required affine distances X (refer again to Figure 3 for the exposition). Eventually, the guarantee for the volume preservation is achieved by an iterative use of the relation $\mathrm{Vol}(V_{s-1}) = H \cdot \mathrm{Vol}(V_{s-2})/(s-1)$.

Proposition 1. *Let $\varepsilon \leq \frac{1}{12}$, let $S \subset \mathbb{R}^n$ be a set of size $s < k$ points, and let $x \in \mathbb{R}^n - S$. Then there is a subset $W = W_{x,S,\varepsilon}$ of $\mathbb{R}^n$ of size $\leq (5s)^{\frac{3}{2}s}$, such that if $f : \mathbb{R}^n \to \mathbb{R}^t$ is a linear embedding that does not expand distances in W and does not contract them by more than $1+\varepsilon$, then $\mathrm{ad}(x, S)/(1 + 3\varepsilon) \leq \mathrm{ad}(f(x), f(S) \leq \mathrm{ad}(x, S)$.*

Proof. Let $W_0 = \{x, P(x, S), c_1(x, S), \ldots, c_{s-1}(x, S)\}$. Recall by definition that

- $\forall i, c_i(x, S) \in \mathcal{L}(S)$.
- The vectors $\{x - P(x, S), c_1(x, S) - P(x, S), \ldots, c_{s-1}(x, S) - P(x, S)\}$ are orthogonal and are of the same length, namely $\mathrm{ad}(x, S)$.

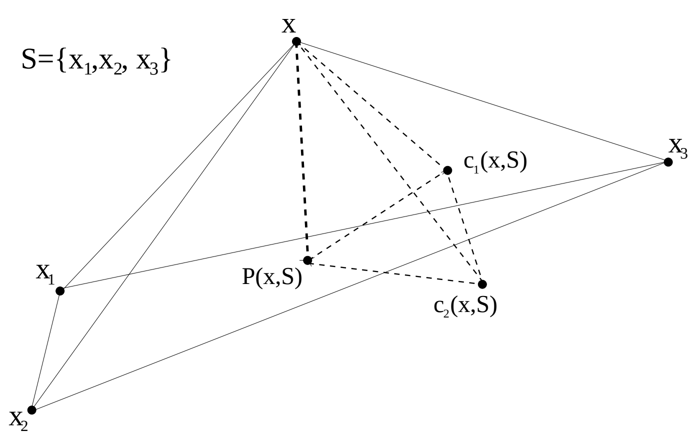

Fig. 3. The set W_0 consists of the vertices of the 'nice' simplex (dashed lines). Notice that the height $[x, P(x, S)]$ is common to the original simplex (solid lines) and to the nice simplex, and that the linear images of these two simplices also share the same height

Clearly, $\mathrm{ad}(x, S) = \mathrm{ad}(x, \{P(x, S), c_1(x, S), \ldots, c_{s-1}(x, S)\}$. Also, by linearity of f, $\mathrm{ad}(f(x), f(S)) = \mathrm{ad}(f(x), \{f(P(x, S)), f(c_1(x, S)), \ldots, f(c_{s-1}(x, S))\})$, and so it is enough to prove that

$$\frac{1}{1+3\varepsilon} \leq \frac{\mathrm{ad}(f(x), f(\{P(x, S), c_1(x, S), \ldots, c_{s-1}(x, S)\}}{\mathrm{ad}(x, \{P(x, S), c_1(x, S), \ldots, c_{s-1}(x, S)\}} \leq 1.$$

Since f is linear, we may simplify and assume that $P(x, S) = 0$ (the zero vector), $\mathrm{ad}(x, S) = 1$ and that $\mathcal{L}(S)$ is spanned by the first $s - 1$ standard vectors (so $c_i(x, S) = e_i$) and also that $x = e_s$. We now need to show that

$$1/(1+3\varepsilon) \leq H \leq 1. \tag{2}$$

where H is the affine distance $\mathrm{ad}(f(e_s), \{0, f(e_1), \ldots, f(e_{s-1})\})$.

The right inequality is immediate, since $H \leq \|f(e_s)\| \leq 1$. We proceed to the more interesting left side of inequality 2. The operation of f on the first s coordinates can be described as a $t \times s$ matrix U. Now, the volume of the the simplex P which is the convex hull of $0, f(e_1), \ldots, f(e_s)$ is $\sqrt{\det(U^t U)}/(s-1)!$. Let $\hat{P}$ be the facet of P opposite $f(e_s)$. Similarly, $\hat{P}$ has volume $\sqrt{\det(\hat{U}^t \hat{U})}/(s-2)!$, where $\hat{U}$ is obtained be removing the last column of U. Consequently

$$H = (s-1) \cdot \mathrm{Vol}(P)/\mathrm{Vol}(\hat{P}) = \sqrt{\det(U^t U)/\det(\hat{U}^t \hat{U})}.$$

We let $A = U_s^t U_s$ and $B = \hat{U}^t \hat{U}$, and note that $A_{ij} = \langle f(e_i), f(e_j) \rangle$, and that B is the principal minor of A that is obtained by removing its last row and column. Our aim is therefore to bound $\det A/\det B$ from below. At this point we divide our analysis depending on the relation between $1/\varepsilon$ and k. The definition of W is dependent on this relation too.

case 1: $1/\varepsilon \geq 4s$. We start with the following algebraic lemma.

Lemma 3. *Let* $\mu \leq \frac{1}{2(s-1)}$, *and let* A *be a real* $s \times s$ *matrix, such that* [2] $\|A - I\|_\infty \leq \mu$. *Denote by* B *the principal minor as described above; then* $\det(A)/\det(B) \geq 1 - 2\mu$.

The proof of Lemma 3 will be given shortly. We now show that for the case $1/\varepsilon > 4s$ it implies proposition 1. Indeed, we take $W = W_0$. Since f is a contraction with distortion $\leq 1+\varepsilon$ on W, for all i we have that $\sqrt{1-2\varepsilon} \leq 1/(1+\varepsilon) \leq \|f(e_i)\| \leq 1$. Now for $i \neq j$ consider the triangle $e_i, 0, e_j$. The first statement of Lemma 2 says that $|\langle f(e_i), f(e_j)\rangle| \leq 2\varepsilon$. We take A and B to be the matrices described above, and we take $\mu = 2\varepsilon$. We have just established that $\|A - I\|_\infty \leq \mu$, and since $1/\varepsilon \geq 4s$ it follows that $\mu \leq \frac{1}{2(s-1)}$. By the proceeding analysis, $H = \sqrt{\det A/\det B} \geq \sqrt{1-2\mu} = \sqrt{1-4\varepsilon} \geq \frac{1}{1+3\varepsilon}$ for $\varepsilon \leq \frac{1}{12}$. To conclude, notice that $|W| = s+1 \leq (5s)^{\frac{3}{2}s}$. We now prove Lemma 3.

Proof. (Lemma 3) We first observe that if $y \in \mathbb{R}^s$ and $v = By$, then $\|v\|_\infty \geq \frac{1}{2}\|y\|_\infty$: Assume without loss of generality that $|y_1| = \|y\|_\infty$. Now,

$$\begin{aligned} |v\|_\infty \geq |v_1| = |\sum_{i<s} y_i a_{1,i}| &= |y_1 + \sum_{i<s} y_i (a_{1,i} - \delta_{1,i})| \\ \geq |y_1| - \mu \sum_{i<s} |y_i| \geq |y_1| - (s-1)\mu|y_1| &\geq \frac{1}{2}|y_1| = \frac{1}{2}\|y\|_\infty \end{aligned}$$

($\delta_{i,j}$ denotes the Kronecker Delta, which is 1 if $i = j$ and 0 otherwise). One immediate consequence of the above is that B is nonsingular, since it implies that if v is the zero vector then so is y.

Now, let v be the vector $(a_{1,s}, a_{2,s}, \ldots, a_{s-1,s})$. For every $j < s$, let $B^{(j)}$ be the matrix B with the j-th column replaced by v. We next argue that $|\frac{\det(B^{(j)})}{\det(B)}| \leq 2\mu$ for all $1 \leq j < s$. Let $y_j = \frac{\det(B^{(j)})}{\det(B)}$. Recall that by Cramer's rule, $y = (y_1, y_2, \ldots, y_{s-1})$ is the (unique) solution to $By = v$. We now get $|\frac{\det(B^{(j)})}{\det(B)}| = |y_j| \leq \|y\|_\infty \leq 2\|v\|_\infty = 2\max_{i<s} a_{i,s} \leq 2\mu$.

Now,

$$\det(A) = a_{s,s}\det(B) + \sum_{i<s}(-1)^{i+s} a_{s,i}\det(B^{(i)}).$$

Therefore

$$\frac{\det(A)}{\det(B)} = a_{s,s} + \sum_{i<s}(-1)^{i+s} a_{s,i}\frac{\det(B^{(i)})}{\det(B)} \geq 1 - \mu - (s-1)\mu \cdot 2\mu \geq 1 - 2\mu.$$

case 2: $1/\varepsilon < 4s$. In this case W_0 alone is too sparse to provide us with the needed bound on the contraction of H. The approach we take here follows an argument used by Feige in [Fei00][3]: Instead of taking just W_0

[2] $\|M\|_\infty$ denote the maximal absolute value of the elements of the matrix M.

[3] although considerably modified for the present use

we add to it a much denser set, namely an $O(\eta)$-net in the unit ball, where $\eta = \varepsilon/\sqrt{s}$. Such a net in the present case where $1/\varepsilon < 4s$ is not too big as a function of s, but still good enough for the bound on the contraction of H.

We turn to the details of this construction. Call a set η-*separated* if the distance between any two distinct points in the set is greater than η. We take W to be an inclusion-maximal η-separated subset of the unit ball. By standard arguments, for any point v in the unit ball there is a point $v' \in W$ such that $\|v - v'\| \leq \eta$ (otherwise, $W \bigcup \{v\}$ would also be a η-separated subset of the ball). Without loss of generality we can assume that $W \supset W_0$. We now assume that f does not contract distances in W by more than $1+\varepsilon$. We first bound the norm of U (as a linear operator). Since f is a contraction on W_0, the columns of U are of (Euclidean) norm at most 1. This means that $\|U\|_2 = \max\{\|Uv\| : \|v\| = 1\} \leq \sqrt{s}$. We next show that if v is a unit vector then $\|Uv\| \geq 1 - 3\varepsilon$. Indeed, let v' be a vector in W such that $\|v - v'\| \leq \eta$. Now

$$\|Uv\| = \|Uv' + U(v - v')\| \geq \|Uv'\| - \|U(v - v')\| \geq \frac{1}{1+\varepsilon} - \sqrt{s}\|v - v'\| \geq$$

$$\frac{1}{1+\varepsilon} - \eta\sqrt{s} = \frac{1}{1+\varepsilon} - \sqrt{s} \cdot \frac{\varepsilon}{\sqrt{s}} \geq 1 - 2\varepsilon.$$

Now, let $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_s$ be the eigenvalues of A, and let $\sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_{s-1}$ be the eigenvalues of B. It is known that $\lambda_i \leq \sigma_i \leq \lambda_{i+1}$, and so

$$\frac{\det A}{\det B} = \frac{\prod_{i=1}^{s} \lambda_i}{\prod_{i=1}^{s-1} \sigma_i} \geq \lambda_1.$$

We now use the standard fact that λ_1, the minimal eigenvalue of A, is $(\min\{\|Uv\| : \|v\| = 1\})^2$. It follows that

$$H = \sqrt{\det A / \det B} \geq \sqrt{\lambda_1} = \min\{\|Uv\| : \|v\| = 1\} \geq 1 - 2\varepsilon \geq 1/(1+3\varepsilon)$$

for $\varepsilon < 1/12$. It remains to show that W is not too big. Since W is η-separated, by a volume argument $|W| \leq (\frac{2+\eta}{\eta})^s$. Therefore $|W| \leq (\frac{2+\eta}{\eta})^s = (1 + \frac{2\sqrt{s}}{\varepsilon})^s \leq (5s)^{\frac{3}{2}s}$.

Corollary 1. *Let $\varepsilon \leq \frac{1}{4}$, let $S \subset \mathbb{R}^n$ be a set of size k points, and let $x \in \mathbb{R}^n - S$. Let f be a random projection onto t dimension multiplied by $\sqrt{\frac{N}{t(1+\varepsilon)}}$. Then*

$$\text{Prob}\left(\text{ad}(x, S)/(1+\varepsilon) \leq \text{ad}(f(x), f(S)) \leq \text{ad}(x, S)\right) \geq 1 - \exp(3k(2 + \log k) - \frac{2}{135}t\varepsilon^2)$$

Proof. Let $W = W_{x,S,\varepsilon/3}$. By Lemma 1 we get that the probability that f is a contraction with distortion $\leq 1+\varepsilon/3$ on W is at least $1 - \binom{|W|}{2} \exp(-\frac{2}{15}t(\frac{\varepsilon}{3})^2)$. By proposition 1 such an embedding satisfies $\text{ad}(x, S)/(1+\varepsilon) \leq \text{ad}(f(x), f(S)) \leq \text{ad}(x, S)$. Now $|W| \leq (5k)^{\frac{3}{2}k}$ and so $\binom{|W|}{2} \leq |W|^2 \leq (5k)^{3k} \leq \exp(3k(2 + \log k))$, and the bound in the lemma follows.

Remark 2. The different approaches we use for the two cases in Proposition 1 seem to be necessary, in the sense that no one of them can be applied to the other case: suppose we apply the approach of case 1 when $\varepsilon = \frac{1}{2s}$, then the linear embedding $f(e_j) = e_j - \frac{1}{s}\sum_i f(e_i)$ has distortion $\leq 1+\varepsilon$, but in the same time makes H zero!. If, on the other hand, we use a dense net when $1/\varepsilon \gg n$ it means that $|W| \gg n^s$, which in turn leads to a dimensionality $O(\varepsilon^{-2}k\log(1/\varepsilon))$ rather than $O(\varepsilon^{-2}k\log n)$.

4.1 The Main Theorem

Theorem 2. *Let $\varepsilon \leq \frac{1}{4}$ and let k, n, t be integers greater than 1, for which $t \geq 70\varepsilon^{-2}(k\log n + 3k(2+\log k))$. Then for any n-point subset X of the Euclidean space $\mathbb{R}^n$, there is a linear mapping $f : X \to \mathbb{R}^t$, such that for all subsets S of X, $1 < |S| < k$,*

$$\text{Vol}(S)/(1+\varepsilon) \leq \text{Vol}(f(S)) \leq \text{Vol}(S),$$

and for $x \in X - S$,

$$\text{ad}(x, S)/(1+\varepsilon) \leq \text{ad}(f(x), f(S)) \leq \text{ad}(x, S).$$

Proof. We apply Corollary 1 to all choices of S and x as above, and then use union bound. We get that as long as

$$\sum_{s=2}^{k} s\binom{n}{s} \exp\left(3k(2+\log k) - \frac{2}{135}t\varepsilon^2\right) < 1,$$

the probability that a random projection onto t dimensions is a contraction on X and all relevant affine distances are preserved to within $1+\varepsilon$, is positive. Now

$$\sum_{s=2}^{k} s\binom{n}{s} \exp\left(3k(2+\log k) - \frac{2}{135}t\varepsilon^2\right) \le n^k \exp\left(3k(2+\log k) - \frac{2}{135}t\varepsilon^2\right) =$$

$$\exp\left(k\log n + 3k(2+\log k) - \frac{2}{135}t\varepsilon^2\right).$$

Setting $t = 70\varepsilon^{-2}(k\log n + 3k(2+\log k)) = O(\varepsilon^{-2}k\log n)$ then, guarantees that with positive probability f preserves all affine distance of sets of size at most k to within relative error of ε.

We turn to the other part of the theorem, namely volume preservation. Let $S_r = x_1, x_2, \ldots, x_r$. It is known that $\mathrm{Vol}(S_r) = \frac{1}{(r-1)!} \cdot \prod_{i=1}^{r-1} \mathrm{ad}(x_{i+1}, S_i)$, and so the volume distortion of S, $(\mathrm{Vol}(S)/\mathrm{Vol}(f(S)))^{\frac{1}{r-1}}$ is simply the geometric mean of $\{\mathrm{ad}(x_{i+1}, S_i)/\mathrm{ad}(f(x_{i+1}), f(S_i))\}_{i=1}^{r-1}$. We now conclude $1 \le (\mathrm{Vol}(S)/\mathrm{Vol}(f(S)))^{\frac{1}{r-1}} \le 1+\varepsilon$.

5 Applications

5.1 The Approximated Nearest Neighbor to Affine Spaces problem

Let $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ be k-dimensional flats in $\mathbb{R}^d$, and let $x \in \mathbb{R}^d$ be a query point. To answer a Nearest Neighbor to Affine-Spaces is to find the flat closest (in the Euclidean sense) to x. In [Mei93] Meiser presents a solution to the point location problem in arrangements of n hyperplanes in $\mathbb{R}^d$ with running time $O(d^5 \log n)$ and space $O(n^{d+1})$. That paper was a breakthrough in that it was the first time (and to the best of our knowledge also the last) where an algorithm to the problem was presented, that is not exponential neither in d nor in $\log n$. We show that by combining this result with ours, a considerable improvement to the problem of approximating the Nearest Affine Neighbor can be achieved. Let $\rho_i : \mathbb{R}^d \to \mathbb{R}$ be the functions that are the squares of distances from the flats, i.e. $\rho_i(x) = dist(x, \mathcal{F}_i)$. Note that ρ_i^2 are polynomials of degree 2 in $x_1, \ldots, x_d$. We now use the following standard linearization of such polynomials. We use the transformation $\xi : \mathbb{R}^d \to \mathbb{R}^{d'}$ which is the assignment transformation to all the monomials of degree at most two in $x_1, \ldots, x_d$. For example, if $d = 2$, $\xi(x) = \xi((x_1, x_2)) = (1, x_1, x_2, x_1x_2)$. Clearly, $d' = \binom{d}{2} + 2d + 1$. Next, we introduce the functions ρ_i' defined by $\rho_i'(\xi(x)) = \rho_i^2(x)$. It is easy to see now that the ρ_i' are linear functionals [4]. We can now reduce the *Nearest Affine Neighbor* problem with query point x to the Vertical Ray Shooting problem from x: 'Shoot' a ray from $(\xi(x), -\infty)$ 'upwards', in other words to the positive direction of the last coordinate. Let $\mathcal{S}_i$ be (one of) the first surface this ray hits. Then $\mathcal{F}_i$ is the closest of $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_s$ to x. This vertical ray-shooting query in linear arrangements can be easily answered using the data-structure of Meiser.

The Preprocessing time, as well as space needed for the above are $O(n^{d'+1}) = O(n^{d^2})$, and the query time is $O(d'^5 \log n) = O(d^{10} \log n)$. We now apply our result in the following way.

For each flat $\mathcal{F}_i$, take an affinely-independent subset of $k+1$ points. We can embed these $O(kn)$ points to $O(\varepsilon^{-2}k\log(kn)) = O(\varepsilon^{-2}k\log n)$ dimensions, and for k constant this is just $O(\varepsilon^{-2}\log n)$. Concatenation of the above yields a query time of $O(\varepsilon^{-20}\log^{11} n)$, and a preprocessing time (and space) of $n^{O(\varepsilon^{-4}\log^2 n)}$.

The standard problem in this approach remains: we need to answer queries on any point of the space, and not on a predetermined set. Looking into the proof of Theorem 2 we easily notice that by taking $t = \varepsilon^{-2}Qk\log n$ we get that the probability that a fixed projection is good in the sense that it does not distort affine distances from a random point x on the sphere to sets of size at most $k-1$ in X is $n^{-\Omega(Q)}$. We can therefore approximate the answer to all but a small fraction of the query points, with only a constant factor sacrifice in the number of dimensions needed.

[4] The ρ_i' are no longer defined on the whole space, but this should not be of any concern to us.

5.2 Projective Clustering

Projective Clustering is a well known problem which has important applications in data-mining. This special variant of clustering, relates very closely to the geometric structures that are discussed in this paper. Here is the problem: Given an n-point set X in $\mathbb{R}^N$ and an integer $s > 0$, find s k-dimensional flats (k-flats), $h_1, \ldots, h_k$, so that the greatest distance of any point in X from its closest flat is minimized. In other words, this is a s-clustering problem, where a cluster is defined by a k-flat, and the quality of the clustering, its *width*, is the maximal distance of any point from its corresponding flat [5]. The "regular" clustering is in fact a special case, where $k = 0$. There are different ways to define the solution for the *projective clustering* problem: the optimal width, the clustering, and the k-flats themselves. In [PC00], Agarwal and Procopiuc give an efficient approximation algorithm for the planar case where the flats are lines, i.e. $N = 2$ and $k = 1$. In [HV02], Har-Peled and Varadarajan give a $dn^{O(\frac{ks}{\varepsilon^5}\log(1/\varepsilon))}$ algorithm that approximates the solution to within $1+\varepsilon$. Here we show that under certain restrictions one can reduce the problem to one in which the space is of dimension that depends logarithmically in n, such that only a small inaccuracy incurs.

In [PC00], the original problem is reduced to the variant where the candidate flats are only the X-k-flats. Namely only affine-subspaces that are spanned by $k+1$ points from X are considered. We call this variant the *Median Projective Clustering*. Agarwal et al. show that by this reduction an additional factor of at most 2 is added to the approximation. We claim

Theorem 3. *An instance of the* Median Projective Clustering *with $X \subseteq \mathbb{R}^N$ and k the dimensionality of the flats, can be reduced to a t-dimensional space instance with a $1+\varepsilon$ approximation to the optimal width, where $t = O(\varepsilon^{-2} k \log n)$,.*

Proof. Using Theorem 2, we can map X to $\mathbb{R}^t$, such that the distance of points in X to k-dimensional X-flats do not expand, and do not contract by more than a factor of $1+\varepsilon$ Therefore any solution w^* in the reduced problem corresponds to a solution w in the original problem, with $w^* \leq w \leq 2w^*$, and we immediately get that $\tau^* \leq \tau \leq 2\tau^*$, where τ, τ^* are the optimal solution to the original problem, and the optimal solution to the reduced problem respectively. Note that the clustering for that width can also be given from the reduced problem. It is not clear, however, how to reconstruct the k flats from the ones in the reduced problem, as the k-flats in the smaller space do not map back to the original space (a projection is not injective).

Acknowledgments I thank Robert Krauthgamer and Nati Linial for discussions that led to the question of this paper, to Uri Feige for his suggestion to use a dense net in case 2 of Proposition 1, to Piotr Indyk and Sariel Har-Peled for the Computational-Geometry relevant Background and references, and many suggestions for Applications. In particular to Sariel, who pointed out the way to use Meiser's result in Section 5.1.

References

[Ach01] Dimitris Achlioptas. Database-friendly random projections. In *roceedings of PODS 01*, pages 274–281, 2001.

[AV99] R.I. Arriaga and S. Vempala. An algorithmic theory of learning: Robust concepts and random projections. In *Proceeding of the 31th Symposium on the Theory of Computing*, New York, 1999.

[DG99] S. Dasgupta and A. Gupta. An elementary proof of the johnson-lindenstrauss lemma. In *Technical Report TR-99-06*, Computer Scinece Institute, Berkeley, CA, 1999.

[EIO02] L. Engebretsen, P. Indyk, and R. O'Donnell. Derandomized dimensionality reduction with applications. In *Proceedings of the 13th Symposium on Discrete Algorithms*. IEEE, 2002.

[Fei00] U. Feige. Approximating the bandwidth via volume respecting embeddings. *J. Comput. System Sci.*, 60(3):510–539, 2000.

[FM88] P. Frankl and H. Mahera. The johnson lindenstrauss lemma and the sphericity of some graphs. *J. Combin. Theory Ser. B*, 44(2):355–362, 1988.

[HV02] S. Har-Peled and K. R. Varadarajan. Projective clustering in high dimensions using core-sets. In *Proc. 18th Annu. ACM Sympos. Comput. Geom.*, pages 312–318, 2002.

[IM98] P. Indyk and R. Motwani. Approximate nearest neighbor: towards removing the curse of dimensionality. In *Proceedings of the Thirty Second Symposium on Theory of Computing*, pages 604–613. ACM, 1998.

[5] Any other variant, such as average distance, sum and sum-of-squares is probably as applicable here.

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[LLR95] N. Linial, E. London, and Yu. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[Mei93] S. Meiser. Point location in arrangements of hyperplanes. *Information and Computation*, 106:286–303, 1993.

[PC00] Agarwal P. and Procopiuc C. Approximation algortihms for strip cover in the plane. In *Proceeding of the 11th ACM-SIAM Symposium on Discrete Algorithms*, pages 373–382, 2000.

[Rao99] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry (Miami Beach, FL, 1999)*, pages 300–306 (electronic), New York, 1999. ACM.

[Vem98] S. Vempala. Approximating vlsi layout problems. In *Proceedings of the thirty ninth Symposium on Foundations of Computer Science*. IEEE, 1998.