

Approximate range searching in higher dimension [☆]

Bernard Chazelle ^a, Ding Liu ^a, Avner Magen ^{b,*}

^a Department of Computer Science, Princeton University, USA

^b Department of Computer Science, University of Toronto, Canada

Received 7 January 2005; received in revised form 10 November 2006; accepted 7 May 2007

Available online 3 July 2007

Communicated by P. Bose and T. Fevens

Abstract

Applying standard dimensionality reduction techniques, we show how to perform approximate range searching in higher dimension while avoiding the curse of dimensionality. Given n points in a unit ball in \mathbb{R}^d , an approximate halfspace range query counts (or reports) the points in a query halfspace; the qualifier “approximate” indicates that points within distance ε of the boundary of the halfspace might be misclassified. Allowing errors near the boundary has a dramatic effect on the complexity of the problem. We give a solution with $\tilde{O}(d/\varepsilon^2)$ query time and $dn^{O(\varepsilon^{-2})}$ storage. For an exact solution with comparable query time, one needs roughly $\Omega(n^d)$ storage. In other words, an approximate answer to a range query lowers the storage requirement from exponential to polynomial. We generalize our solution to polytope/ball range searching.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Approximate range searching; Dimensionality reduction

1. Introduction

A staple of computational geometry [1,2], range searching is the problem of preprocessing a set P of n points in \mathbb{R}^d so that, given a region R (the range) chosen from a predetermined class (e.g., all d -dimensional boxes, simplices, or halfspaces), the points of $P \cap R$ can be counted or reported quickly. The case of halfspaces is noteworthy because many range searching problems with “algebraic” ranges can be reduced to it through linearization-via-lifting. The counting version can be solved in $O(\log n)$ query time and $O(n^d / \log^d n)$ storage, while the reporting case can be handled in $O(\log n + k)$ query time and $O(n^{\lfloor d/2 \rfloor} \text{polylog}(n))$ storage, where k is the number of points to be reported [1]. In both cases, the exponential dependency on d —the so-called curse of dimensionality—is a show-stopper for large d . Lower bound work in a variety of highly reasonable models suggests that the curse of dimensionality is inevitable [4,5].

Inspired by recent work on approximate nearest neighbor searching [8–10], we seek the mildest relaxation of the problem that will break the curse of dimensionality. Without loss of generality we assume that all the points of P lie in a unit ball of ℓ_2^d . Let S_h be a halfspace, with h denoting its bounding hyperplane. Given $\varepsilon > 0$, the *fuzzy boundary*

[☆] This work was supported in part by NSF grants CCR-998817, 0306283, ARO Grant DAAH04-96-1-0181.

* Corresponding author.

E-mail addresses: chazelle@cs.princeton.edu (B. Chazelle), dingliu@cs.princeton.edu (D. Liu), avner@cs.toronto.edu (A. Magen).

of S_h is the slab formed by all points within distance ε of h (Fig. 1). Approximate halfspace range searching refers to counting (or reporting) the points of $P \cap S_h$, making allowance for errors regarding the points in the fuzzy boundary; in other words, the output should be the size of a set whose symmetric difference with $P \cap S_h$ lies entirely in the fuzzy boundary.

Approximate range searching is relevant in situations where the data is inherently imprecise and points near the boundary cannot be classified as being inside or outside with any certainty. In the case of reporting, of course, one can always move the boundary by ε to ensure that the output contains every point of $P \cap S_h$, which then allows us to retrieve the right points by filtering out the outsiders.

Theorem 1.1. *Approximate halfspace range searching can be solved with query time $\tilde{O}(d/\varepsilon^2)$ and $dn^{O(\varepsilon^{-2})}$ storage.¹ Any given query is answered correctly with arbitrarily high probability.*

Our algorithm beats the lower bound for the exact version of the problem. Indeed, it is known that in the arithmetic model if only $dn^{O(\varepsilon^{-2})}$ storage is available, then the query time must be $\Omega(n^{1-O(1/d\varepsilon^2)})$ [4]. Our algorithm generalizes to ranges formed by polytopes bounded by a fixed number of hyperplanes and to (Euclidean) ball range searching.

We also propose an alternative algorithm for approximate halfspace range searching with a query time of $\tilde{O}((d/\varepsilon)^2 + dn^{1/(1+\varepsilon)}\varepsilon^{-2})$ and storage $\tilde{O}(dn\varepsilon^{-2} + n^{1+1/(1+\varepsilon)})$ —and a slightly different definition of approximation. Again, the query time is better than the solution for the exact problem, since by [4] $\Omega(n^{1-O(1)/d})$ query time is necessary when we have close to $O(n^2)$ space.

Approximate range searching does not originate with this paper. Arya and Mount [3] gave an algorithm for the problem that uses optimal $O(dn)$ storage but provides a query time of $O(\log n + \varepsilon^{-d})$, which is exponential in d . Their algorithm is based on space partition techniques, which is very different from our dimension reduction approach.

2. Approximate halfspace range searching

In this section we show how to reduce approximate halfspace range searching to an approximate variant of ball range searching in the Hamming cube. Initially, we make the “homogeneous” assumption that the hyperplanes bounding the query halfspaces pass through the origin and that all of the n points lie in the Euclidean ball $\|x\|_2 \leq 1$. We relax the homogeneous condition later by lifting to one dimension higher.

2.1. The homogeneous case

Let v_h be the unit vector normal to h pointing inside S_h . Any point p_1 in S_h outside the fuzzy boundary is at a distance from h at least ε (Fig. 1). It follows that the angle between Op_1 and v_h is less than $\pi/2 - \varepsilon$. (We assume that $\varepsilon <$

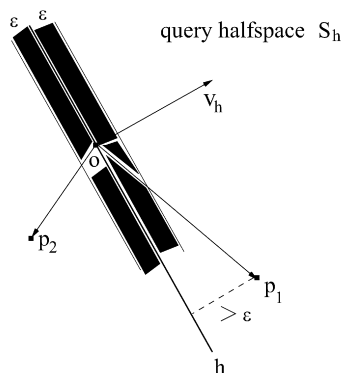


Fig. 1. Approximate halfspace range searching.

¹ The notation $\tilde{O}(f)$ stands for $O(f \text{polylog}(f \cdot n))$.

$\pi/2$.) Similarly, for a point p_2 not in S_h and outside the fuzzy boundary, the angle between Op_2 and v_h is greater than $\pi/2 + \varepsilon$. This provides a separation criterion to distinguish between points we must include and those we must not.

Let S^{d-1} denote the unit $(d-1)$ -sphere in \mathbb{R}^d and let $\text{sign}(t)$ be 1 if $t \geq 0$ and -1 otherwise. Let x and y be two vectors in \mathbb{R}^d and let $0 \leq \theta_{x,y} \leq \pi$ be the angle between them. We use $\mathcal{E}_{x,y}$ to denote the event: $\text{sign}(x \cdot u) = \text{sign}(y \cdot u)$. If u is uniformly distributed over S^{d-1} , then it is well known that $\text{Prob}[\mathcal{E}_{x,y}] = 1 - \theta_{x,y}/\pi$. It follows that $\text{Prob}[\mathcal{E}_{Op_1, v_h}] > 1/2 + \varepsilon/\pi$ and $\text{Prob}[\mathcal{E}_{Op_2, v_h}] < 1/2 - \varepsilon/\pi$.

Following Kleinberg's approach [9] to nearest neighbor searching, we invoke VC-dimension theory [5,11] to show the existence of a small number of unit vectors that can be used to distinguish between p_1 and p_2 . Let $\mathcal{W}_{x,y}$ denote the subset of S^{d-1} for which $\mathcal{E}_{x,y}$ happens. Let \mathcal{R} be the collection of $\mathcal{W}_{x,y}$, for all $x, y \in \mathbb{R}^d$. We consider the range space (S^{d-1}, \mathcal{R}) . Each range $\mathcal{W}_{x,y}$ is a Boolean combination of four halfspaces; therefore the exponent of its (primal) shatter function is $2d+2$. A finite subset A of S^{d-1} is said to be a γ -approximation for the range space (S^{d-1}, \mathcal{R}) if, for all $R \in \mathcal{R}$, $|\mu(R \cap A)/|A| - \mu(R)| \leq \gamma$. Here $\mu(R)$ is the measure of R in the space S^{d-1} . It follows from VC dimension theory [5] that the range space (S^{d-1}, \mathcal{R}) admits of an $(\varepsilon/(2\pi))$ -approximation A of size $O(d\varepsilon^{-2} \log(d\varepsilon^{-1}))$. Moreover, a randomly chosen set A of that size is good with high probability.

Thus, $|\mathcal{W}_{v_h, Op_1} \cap A|/|A| \geq \mu(\mathcal{W}_{v_h, Op_1}) - \varepsilon/(2\pi) > 1/2 + \varepsilon/(2\pi)$. Similarly, $|\mathcal{W}_{v_h, Op_2} \cap A|/|A| < 1/2 - \varepsilon/(2\pi)$. For any vector x let $\hat{x} \in \{+1, -1\}^{|A|}$ be defined as follows: the i th coordinate of \hat{x} is $\text{sign}(x \cdot u_i)$, where u_i is the i th vector in A according to a fixed ordering. Recall that $|\mathcal{W}_{x,y} \cap A|$ is the number of vectors $u \in A$ such that $\text{sign}(x \cdot u) = \text{sign}(y \cdot u)$. So $|\mathcal{W}_{x,y} \cap A| = |A| - d_H(\hat{x}, \hat{y})$ where $d_H(\cdot, \cdot)$ is the Hamming distance. We thus have $d_H(\hat{v}_h, \widehat{Op_1}) < (1/2 - \varepsilon/(2\pi))|A|$ and $d_H(\hat{v}_h, \widehat{Op_2}) > (1/2 + \varepsilon/(2\pi))|A|$.

It immediately follows that approximate halfspace range searching (under the homogeneous condition) reduces to *approximate ball range searching in the Hamming cube*: Preprocess n points in $\{+1, -1\}^{|A|}$ so that, given any \hat{v}_h , the points in the Hamming ball centered at \hat{v}_h with radius $|A|/2$ can be approximately counted (or reported) quickly. The term "approximately" means that all points within distance $(1/2 - \varepsilon/(2\pi))|A|$ must be included while all points further than $(1/2 + \varepsilon/(2\pi))|A|$ must be excluded.

2.2. The general case

To remove the homogeneous condition, we lift the problem into \mathbb{R}^{d+1} . First of all, we may assume that the input point set lies in the Euclidean ball $\|x\|_2 \leq 1$. Now, map each point $p = (p_1, \dots, p_d)$ to $p' = (p_1, \dots, p_d, 1) \in \mathbb{R}^{d+1}$. Note that the new point set in \mathbb{R}^{d+1} lies in the Euclidean ball $\|x\|_2 \leq \sqrt{2}$. Given a query halfspace: $q_1x_1 + \dots + q_dx_d \geq q_{d+1}$, first we compute the distance from O to its bounding hyperplane. If it exceeds 1, then all of the n points are on one side of the hyperplane and we return the exact answer (either 0 or n). Otherwise, $q_{d+1}^2 \leq \sum_i q_i^2$. We map the query to a new halfspace in \mathbb{R}^{d+1} : $q_1x_1 + \dots + q_dx_d - q_{d+1}x_{d+1} \geq 0$. Note that the new query passes through the origin. Moreover, it is straightforward to verify that: (i) all point-halfspace incidence relations are preserved by the map; and (ii) point-hyperplane distances are preserved to within a factor of $\sqrt{2}$ because of the upper bound on q_{d+1}^2 . The problem is now reduced to the homogeneous case after suitable rescaling of ε .

3. Approximate ball range searching in the Hamming cube

We give two solutions for approximate ball range searching in the k -dimensional Hamming cube H^k , where $k = |A| = O(d\varepsilon^{-2} \log(d\varepsilon^{-1}))$. Recall that the problem is to preprocess a set S of n points so that, given any $q \in H^k$, we can quickly count (or report) approximately the points of S within distance $k/2$ to q .

3.1. A high-storage solution

We adapt to the problem at hand Kushilevitz et al.'s solution to approximate nearest neighbor searching in the Hamming cube [10]. Fix two parameters m and t to be determined later. The search structure \mathcal{S} consists of m substructures $\mathcal{T}_1, \dots, \mathcal{T}_m$, all of them constructed in the same way but independently of one another. Fix $i \in \{1, \dots, m\}$. The substructure \mathcal{T}_i is built by picking t coordinates of H^k at random (out of k). Project each point $x \in H^k$ onto the subspace spanned by these t coordinates. The resulting vector $t_i(x) \in \{0, 1\}^t$ is called the *trace* of x . Each \mathcal{T}_i consists of a table of 2^t entries, one for each trace. Each entry stores a number (for range counting) or a pointer to a list of points (for range reporting), to be specified below. The intuition is that, as long as t is large enough, say $t = \Theta(\varepsilon^{-2} \log n)$, by a

discrete analogue of Johnson and Lindenstrauss’s theorem [7], the random projections preserve inter-point distances in appropriate range within a relative error of ε .

We say that a substructure \mathcal{T}_i fails at query $q \in H^k$ if there exists $p \in S$ such that either of the following holds:

- $d_H(p, q) < (1/2 - \varepsilon/(2\pi))k$ but $d_H(t(p), t(q)) > (1/2 - \varepsilon/(3\pi))t$;
- $d_H(p, q) > (1/2 + \varepsilon/(2\pi))k$ but $d_H(t(p), t(q)) < (1/2 + \varepsilon/(3\pi))t$.

Lemma 3.1. *The probability that \mathcal{T}_i fails at q is at most $ne^{-\Omega(\varepsilon^2 t)}$.*

Let $0 < c < 1$ be a constant to be specified later. We say that the structure \mathcal{S} fails at q if more than cm sub-structures \mathcal{T}_i fail at q .

Lemma 3.2. *For any $\gamma > 0$ if we set $m = (k + \log \gamma^{-1})/c$, then for some $t = O(\varepsilon^{-2} \ln(2en/c))$, \mathcal{S} fails nowhere with probability at least $1 - \gamma$.*

The proofs of Lemmas 3.1 and 3.2 follow from standard applications of the Chernoff bounds and can be found in [10]. We don’t repeat them here. Lemma 3.2 implies that, with high probability, for any query $q \in H^k$ and any $p \in S$, there are at least $(1 - c)m$ substructures \mathcal{T}_i that provide the following guarantees: (i) if $d_H(p, q) < (1/2 - \varepsilon/(2\pi))k$ then $d_H(t(p), t(q)) \leq (1/2 - \varepsilon/(3\pi))t$; (ii) if $d_H(p, q) > (1/2 + \varepsilon/(2\pi))k$ then $d_H(t(p), t(q)) \geq (1/2 + \varepsilon/(3\pi))t$. In the preprocessing stage, for each entry $t_i(x)$ in the table associated with \mathcal{T}_i , we store the number of points $p \in S$ such that $d_H(t_i(x), t_i(p)) \leq (1/2 - \varepsilon/(3\pi))t$ (for range reporting, we store a pointer to a list of such points). To answer a query q , we pick one substructure $\mathcal{T}_i \in \mathcal{S}$ uniformly at random. We compute $t_i(q)$ and use it to index the table of \mathcal{T}_i . We output the answer stored at that entry. By Lemma 3.2, with probability at least $1 - c$, the substructure \mathcal{T}_i does not fail at q , and so we get a correct answer for approximate ball range queries. It is easy to see that the storage requirement is $\tilde{O}(nd + m2^t)$ in terms of the number of bits: $O(nd)$ for the set of n input points; $m2^t \log n$ for $m2^t$ table entries each storing an answer (a number between 0 and n) for some queries; $mt \log k$ for m substructures each defined by t (random) coordinates out of k . For reporting, the last term above is $m2^t n$. The query time is essentially the time needed to compute $t_i(q)$ (plus the output size for reporting). There are t coordinates to be computed and each takes time $O(d)$ by doing an inner product between the query and a vector from the $(\varepsilon/(2\pi))$ -approximation in Section 2.1. So the query time is $O(dt)$. In view of Lemma 3.2 and the reduction shown in the last section, this proves Theorem 1.1.

We claim that the above algorithm, after some suitable modification, also works when each query is the intersection of a set of halfspaces. For a halfspace S_h , we use S_h^ε to denote its fuzzy boundary. We use $S_h^- = S_h \setminus S_h^\varepsilon$ to denote the part of S_h outside the fuzzy boundary. Similarly, we use $S_h^+ = \overline{S_h} \setminus S_h^\varepsilon$ to denote the part of $\overline{S_h}$ (the complement of S_h) outside the fuzzy boundary. Given a set H of l halfspaces with common intersection Q , we define $Q^- = \bigcap_{S_h \in H} S_h^-$ and $Q^+ = \bigcup_{S_h \in H} S_h^+$. We require that all points in Q^- must be included while all points in Q^+ must be excluded. The region bounded by ∂Q^+ and ∂Q^- is the query’s fuzzy boundary. Unlike the convention in [3], the width of the fuzzy boundary here is not related to the size of the query range but to the diameter of the point set. Using the reduction in the previous section, this problem is reduced to the following multiple-ball approximate range searching in the Hamming cube: Preprocess a set of n points in H^k so that, given any q_1, \dots, q_l all in H^k , we can quickly count (or report) approximately the points within distance $k/2$ to each q_j ($1 \leq j \leq l$). The term “approximately” means that all points within distance $(1/2 - \varepsilon/(2\pi))k$ to every q_j must be included, while all points further than $(1/2 + \varepsilon/(2\pi))k$ from at least one q_j must be excluded.

It is straightforward to modify the algorithm in this section to incorporate such multiple-ball queries. For example, we combine l traces $(t_i(q_1), \dots, t_i(q_l))$ of \mathcal{T}_i together to form a single vector in dimension lt . We also modify each \mathcal{T}_i so that its table has 2^{lt} entries, one for each possible l -tuple of traces. For each entry $(t_i(q_1), \dots, t_i(q_l))$ in the table, we store the number of points p such that $d_H(t_i(q_j), t_i(p)) \leq (1/2 - \varepsilon/(3\pi))t$ holds for each q_j . Answering a query is the same as before: We pick one substructure \mathcal{T}_i uniformly at random. We compute $(t_i(q_1), \dots, t_i(q_l))$ and use it to index the table of \mathcal{T}_i . We output the answer stored at that entry. The definitions and lemmas in this section only need to be changed slightly for the analysis to work. In particular, the values of m and t in Lemma 3.2 are changed to

$(lk + \log \gamma^{-1})/c$ and $O(\varepsilon^{-2} \ln(2enl/c))$ respectively. As long as the number of halfspaces l is constant, the time and space bounds of Theorem 1.1 remain the same.

3.2. A low-storage solution

The storage achieved in the previous section is polynomial in n but with an exponent of $O(\varepsilon^{-2})$. We propose another solution that uses roughly quadratic space and still provides sublinear query time. For this purpose, however, we need to relax the meaning of approximation further. If $N_r(q)$ denotes the number of points of S in the Hamming ball centered at q of radius r , then we output a number N such that $(1 - \alpha)N_{(1-o(\varepsilon))k/2}(q) \leq N \leq (1 + \alpha)N_{(1+o(\varepsilon))k/2}(q)$, for any fixed $\alpha > 0$. In Section 3.6 of [6], it is shown that computing such a number N can be reduced to the $(1 + \varepsilon)$ -PLEB problem (stands for “Point Location in Equal Balls”) with a multiplicative overhead of $\alpha^{-3} \log^2 n$ in both query time and storage. The $(1 + \varepsilon)$ -PLEB problem is defined as follows [6,8]: given a set P of n points in the Hamming cube H^k and a fixed $r \leq k$, preprocess P such that, given any query $q \in H^k$,

- if there exists a point $p \in P$ such that $d_H(p, q) \leq r$, then answer “yes” and return a point $p' \in P$ such that $d_H(p', q) \leq (1 + \varepsilon)r$;
- if $d_H(p, q) > (1 + \varepsilon)r$ for any $p \in P$ then answer “no”.

It is shown in [8] that $(1 + \varepsilon)$ -PLEB in the Hamming cube H^k can be solved with query time $O(kn^{1/(1+\varepsilon)})$ and storage $(kn + n^{1+1/(1+\varepsilon)})$. Therefore approximate ball range searching can be solved with query time $\tilde{O}(dn^{1/(1+\varepsilon)}\varepsilon^{-2})$ and storage $\tilde{O}(dn\varepsilon^{-2} + n^{1+1/(1+\varepsilon)})$, following the above reduction and $k = O(d\varepsilon^{-2} \log(d\varepsilon^{-1}))$. This leads to an algorithm for approximate halfspace range searching with query time $\tilde{O}(d^2\varepsilon^{-2} + dn^{1/(1+\varepsilon)}\varepsilon^{-2})$ and storage $\tilde{O}(dn\varepsilon^{-2} + n^{1+1/(1+\varepsilon)})$, as claimed in the introduction.

4. Approximate ball range searching in Euclidean space

Another problem we can solve is approximate ball range searching in Euclidean space. Given a ball $B(q, r)$ in \mathbb{R}^d with center q and radius r , approximate ball range searching includes all points inside the smaller ball $B(q, r - s\varepsilon)$ while excluding all points outside the larger ball $B(q, r + s\varepsilon)$, for some parameter $s = s(r)$. Points in the annulus $B(q, r + s\varepsilon) \setminus B(q, r - s\varepsilon)$ may be misclassified. In the Hamming cube, the technique described in previous section solves approximate ball range searching for $s = \Theta(r)$. On the other hand, in such a solution the width of the annulus (the fuzzy region) grows with r . When r is large, it might be too big to provide an estimation of the true answer. We give another solution in which s is bounded even when r is large. Moreover, it works in Euclidean space.

Given n points in \mathbb{R}^d in the unit ball $\|x\|_2 \leq 1$ and a query ball $B(q, r)$, we first compute the distance from $q = (q_1, \dots, q_d)$ to the origin. If this distance is greater than $r + 1$, then the query ball contains no points. So from now on we assume that $\|q\|_2 \leq r + 1$, or,

$$\sum_1^d q_i^2 \leq (r + 1)^2 \tag{1}$$

The query ball is given by the following equation:

$$\sum_1^d 2q_i x_i - \sum_1^d x_i^2 + r^2 - \sum_1^d q_i^2 \geq 0$$

By a standard lifting map, it is mapped to a halfspace $S_{q,r}$ in \mathbb{R}^{d+1} :

$$2q_1 x_1 + \dots + 2q_d x_d - x_{d+1} + \left(r^2 - \sum_1^d q_i^2 \right) \geq 0$$

A point $p = (p_1, \dots, p_d)$ in \mathbb{R}^d is mapped to $p' = (p_1, \dots, p_d, \sum_i p_i^2)$ in \mathbb{R}^{d+1} . Note that the new point set lies in the Euclidean ball $\|x\|_2 \leq \sqrt{2}$. Moreover, it is easy to check that $p \in B(q, r)$ if and only if $p' \in S_{q,r}$. It remains to

show that if p is outside the annulus $B(q, r + s\varepsilon) \setminus B(q, r - s\varepsilon)$ then p' is outside a fuzzy boundary of $S_{q,r}$ with width $\Theta(\varepsilon)$. Given $p \in B(q, r - s\varepsilon)$ we have:

$$r^2 - \sum_i (q_i - p_i)^2 \geq 2rs\varepsilon - s^2\varepsilon^2 \geq rs\varepsilon \quad (2)$$

The distance from p' to the boundary of $S_{q,r}$ is:

$$\frac{r^2 - \sum_i (q_i - p_i)^2}{\sqrt{1 + 4 \sum_i q_i^2}} \geq \frac{rs\varepsilon}{\sqrt{1 + 4(r+1)^2}} \geq \frac{1}{\sqrt{5}} \cdot \frac{rs\varepsilon}{r+1}$$

Note that the first inequality above follows from Eqs. (1) and (2). Using similar arguments, we can show that, for $p \notin B(q, r + s\varepsilon)$, the distance from p' to the boundary of $S_{q,r}$ is at least $(2rs\varepsilon)/(\sqrt{5}(r+1))$. Setting $s = (r+1)/r$ keeps p' outside a fuzzy boundary of width $\Theta(\varepsilon)$ and hence properly classified by the algorithm for halfspace. In this solution $s = O(1)$ when $r = \Omega(1)$, and so the fuzzy region does not grow with r . The time and space bounds are essentially the same as those for approximate halfspace range searching; in particular, the bounds of Theorem 1.1 apply to approximate ball range searching as well.

References

- [1] P.K. Agarwal, J. Erickson, Geometric range searching and its relatives, in: B. Chazelle, J.E. Goodman, R. Pollack (Eds.), *Advances in Discrete and Computational Geometry*, AMS Press, 1999, pp. 1–56.
- [2] P.K. Agarwal, Range searching, in: J.E. Goodman, J. O'Rourke (Eds.), *Handbook of Discrete and Computational Geometry*, second ed., Chapman and Hall, CRC, 2004.
- [3] S. Arya, D.M. Mount, Approximate range searching, *Comput. Geom. Theory Appl.* 17 (2000) 135–163.
- [4] H. Brönnimann, B. Chazelle, J. Pach, How hard is halfspace range searching, *Discrete Comput. Geom.* 10 (1993) 143–155.
- [5] B. Chazelle, *The Discrepancy Method: Randomness and Complexity*, Cambridge University Press, 2000, Chapter 6, paperback version 2001.
- [6] P. Indyk, High-dimensional computational geometry, PhD Thesis, Stanford University, 2000, <http://theory.lcs.mit.edu/~indyk/thesis.ps>.
- [7] P. Indyk, J. Matousek, Low-distortion embeddings of finite metric spaces, in: J.E. Goodman, J. O'Rourke (Eds.), *Handbook of Discrete and Computational Geometry*, second ed., Chapman and Hall, CRC, 2004.
- [8] P. Indyk, R. Motwani, Approximate nearest neighbors: Towards removing the curse of dimensionality, in: *Proc. ACM STOC*, 1998, pp. 604–613.
- [9] J.M. Kleinberg, Two algorithms for nearest-neighbor search in high dimensions, in: *Proc. ACM STOC*, 1997, pp. 599–608.
- [10] E. Kushilevitz, R. Ostrovsky, Y. Rabani, Efficient search for approximate nearest neighbor in high dimensional spaces, *SIAM J. Comput.* 30 (2000) 457–474.
- [11] V.N. Vapnik, A.Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory Probab. Appl.* 16 (1971) 264–280.