Alcom-FT Technical Report Series
ALCOMFT-TR-03-89

Smoothed Motion Complexity *

Valentina Damerow
PaSCo Graduate School and
Heinz Nixdorf Institute, University of Paderborn, D-33102 Paderborn, Germany

Friedhelm Meyer auf der Heide
Heinz Nixdorf Institute, University of Paderborn, D-33102 Paderborn, Germany

Harald Räcke
Heinz Nixdorf Institute, University of Paderborn, D-33102 Paderborn, Germany

Christian Scheideler
Dept. of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA

Christian Sohler
Heinz Nixdorf Institute, University of Paderborn, D-33102 Paderborn, Germany

Abstract

We propose a new complexity measure for movement of objects, the *smoothed motion complexity*. Many applications are based on algorithms dealing with moving objects, but usually data of moving objects is inherently noisy due to measurement errors. Smoothed motion complexity considers this imprecise information and uses *smoothed analysis* [13] to model noisy data. The input is object to slight random perturbation and the *smoothed complexity* is the worst case expected complexity over all inputs w.r.t. the random noise. We think that the usually applied worst case analysis of algorithms dealing with moving objects, e.g., kinetic data structures, often does not reflect the real world behavior and that smoothed motion complexity is much better suited to estimate dynamics.

We illustrate this approach on the problem of maintaining an orthogonal bounding box of a set of n points in $\mathbb{R}^d$ under linear motion. We assume speed vectors and initial positions from $[-1, 1]^d$. The motion complexity is then the number of combinatorial changes to the description of the bounding box. Under perturbation with Gaussian normal noise of deviation σ the smoothed motion complexity is only polylogarithmic: $O(d \cdot (1 + 1/\sigma) \cdot \log n^{3/2})$ and $\Omega(d \cdot \sqrt{\log n})$. We also consider the case when only very little information about the noise distribution is known. We assume that the density function is monotonically increasing on $\mathbb{R}_{\leq 0}$ and monotonically decreasing on $\mathbb{R}_{\geq 0}$ and bounded by some value C. Then the motion complexity is $O(d \cdot (\sqrt{n \log n \cdot C} + \log n))$ and $\Omega(d \cdot \min\{\sqrt[5]{n}/\sigma, n\})$.

Keywords: Randomization, Kinetic Data Structures, Smoothed Analysis

1 Introduction

The task to process a set of continuously moving objects arises in a broad variety of applications, e.g., in mobile ad-hoc networks, traffic control systems, and computer graphics (rendering moving

*The third and the fifth author are partially supported by DFG-Sonderforschungsbereich 376, DFG grant 872/8-1, and the Future and Emerging Technologies program of the EU under contract number IST-1999-14186 (ALCOM-FT).

objects). Therefore, researchers investigated data structures that can be efficiently maintained under continuous motion, e.g., to answer proximity queries [5], maintain a clustering [8], a convex hull [4], or some connectivity information of the moving point set [9]. Within the framework of kinetic data structures the efficiency of such a data structure is analyzed w.r.t. to the worst case number of combinatorial changes in the description of the maintained structure that occur during linear (or low degree algebraic) motion. These changes are called *(external) events*. For example, to maintain the smallest orthogonal bounding box of a point set in $\mathbb{R}^d$ has a unique description at a certain point of time consisting of the $2d$ points that attain the minimum and maximum value in each of the d coordinates. If any such minimum/maximum point changes then an event occurs. We call the worst case number of events w.r.t. the maintainance of a certain structure under linear motion the *worst case motion complexity*.

We introduce an alternative measure for the dynamics of moving data called the *smoothed motion complexity*. Our measure is based on *smoothed analysis*, a hybrid between worst case analysis and average case analysis. Smoothed analysis has been introduced by Spielman and Teng [13] in order to explain the typically good performance of the simplex algorithm on almost every input. It asks for the worst case expected performance over all inputs where the expectation is taken w.r.t. small random noise added to the input. In the context of mobile data this means that both the speed value and the starting position of an input configuration are slightly perturbed by random noise. Thus the smoothed motion complexity is the worst case expected motion complexity over all inputs perturbed in such a way. Smoothed motion complexity is a very natural measure for the dynamics of mobile data since in many applications the exact position of mobile data cannot be determined due to errors caused by physical measurements or fixed precision arithmetic. This is, e.g., the case when the positions of the moving objects are determined via GPS, sensors, and basically in any application involving 'real life' data.

We illustrate our approach on the problem to maintain the smallest orthogonal bounding box of a point set moving in $\mathbb{R}^d$. The bounding box is a fundamental measure for the extend of a point set and it is useful in many applications, e.g., to estimate the sample size in sublinear clustering algorithms [3], in the construction of R-trees, for collision detection, and visibility culling.

1.1 The Problem Statement

We are given a set P of n points in $\mathbb{R}^d$. The position $\mathrm{pos}_i(t)$ of the ith point at time t is given by a linear function of t. Thus we have $\mathrm{pos}_i(t) = s_i \cdot t + p_i$ where p_i is the initial position and s_i the speed. We normalize the speed vectors and initial positions such that $p_i, s_i \in [-1, 1]^d$.

The motion complexity of the problem is the number of combinatorial changes to the set of $2d$ extreme points defining the bounding box. Clearly this motion complexity is $O(d \cdot n)$ in the worst case, 0 in the best case, and $O(d \cdot \log n)$ in the average case. When we consider smoothed motion complexity we add to each coordinate of the speed vector and each coordinate of the initial position an i.i.d. random variable from a certain probability distribution, e.g., Gaussian normal distribution. Then the smoothed motion complexity is the worst case expected complexity over all choices of p_i and s_i.

1.2 Related Work

In [4] Basch et al. introduced *kinetic data structures* (KDS) which is a framework for data structures for moving objects. In KDS the (near) future motion of all objects is known and can be specified by so-called pseudo-algebraic functions of time specified by linear functions or low-degree polynomials.

This specification is called a *flight plan*. The goal is to maintain the description of a combinatorial structure as the objects move according to this flight plan. The flight plan may change from time to time and these updates are reported to the KDS. The efficiency of a KDS is analyzed by comparing the worst case number of internal (events needed to maintain auxiliary data structures) and external events it processed against the worst case number of external events. Using this framework many interesting kinetic data structures have been developed, e.g., for connectivity of discs [7] and rectangles [9], convex hulls [4], proximity problems [5], and collision detection for simple polygons [10]. In [4] the authors developed a KDS to maintain a bounding box of a moving point set in $\mathbb{R}^d$. The number of events these data structures process is $O(n \log n)$ which is close to the worst case motion complexity of $\Theta(n)$. In [1] the authors showed that it is possible to maintain an $(1 + \epsilon)$-approximation of such a bounding box. The advantage of this approach is that the motion complexity of this approximation is only $O(1/\sqrt{\epsilon})$. The average case motion complexity has also been considered in the past. If n particles are drawn independently from the unit square then it has been shown that the expected number of combinatorial changes in the convex hull is $\Theta(\log^2(n))$, in the Voronoi diagram $\Theta(n^{3/2})$ and in the closest pair $\Theta(n)$ [15].

Smoothed analysis has been introduced by Spielman and Teng [13] to explain the polynomial run time of the simplex algorithm on inputs arising in applications. They showed that the smoothed run time of the shadow-vertex simplex algorithm is polynomial in the input size and $1/\sigma$. In many follow-up papers other algorithms and values have been analyzed via smoothed analysis, e.g., the perceptron algorithm [6], condition numbers of matrices [12], quicksort, left-to-right maxima, and shortest paths [2]. Recently, smoothed analysis has been used to show that many existing property testing algorithms can be viewed as sublinear decision algorithms with low smoothed error probability [14]. In [2] the authors analyzed the smoothed number of left-to-right maxima of a sequence of n numbers. We will use the left-to-right maxima problem as an auxiliary problem but we will use a perturbation scheme that fundamentally differs from that analyzed in [2].

1.3 Our Results

Typically, measurement errors are modelled by the Gaussian normal distribution and so we analyze the smoothed complexity w.r.t. Gaussian normally distributed noise with deviation σ. We show that the smoothed motion complexity of a bounding box under Gaussian noise is $O(d \cdot (1+1/\sigma) \cdot \log n^{3/2})$ and $\Omega(d \cdot \sqrt{\log n})$. In order to get a more general result we consider monotone probability distributions, i.e., distributions where the density function f is bounded by some constant C and monotonically increasing on $\mathbb{R}_{\leq 0}$ and monotonically decreasing on $\mathbb{R}_{\geq 0}$. Then the smoothed motion complexity is $O(d \cdot (\sqrt{n \log n \cdot C} + \log n))$. Polynomial smoothed motion complexity is, e.g., attained by the uniform distribution where we obtain a lower bound of $\Omega(d \cdot \min\{\sqrt[5]{n}/\sigma, n\})$.

Note that in the case of speed vectors from some arbitrary range $[-S, S]^d$ instead of $[-1, 1]^d$ the above upper bounds hold if we replace σ by σ/S.

These results make it very unlikely, that in a typical application the worst case bound of $\Theta(d \cdot n)$ is attained. As a consequence, it seems reasonable to analyze KDS's w.r.t. the smoothed motion complexity rather than the worst case motion complexity.

Our upper bounds are obtained by analyzing a related auxiliary problem: the smoothed number of left-to-right maxima in a sequence of n numbers. For this problem we also obtained lower bounds which only can be stated here: in the case of uniform noise we have $\Omega(\sqrt{n/\sigma})$ and in the case of normally distributed noise we can apply the average case bound of $\Omega(\log n)$. These bounds differ only by a factor of $\sqrt{\log n}$ from the corresponding upper bounds. In the second case the bounds are even tight for constant σ. Therefore, we can conclude that our analysis is tight w.r.t. the number of left-to-right maxima. To obtain better results a different approach that does not use left-to-right maxima as an auxiliary problem is necessary.

2 Upper Bounds

To show upper bounds for the number of external events while maintaining the bounding box for a set of moving points we make the following simplifications. We only consider the 1D problem. Since all dimensions are independently from each other an upper or lower bound for the 1D problem can be multiplied by d to yield a bound for the problem in d dimensions.

Further, we assume that the points are ordered by their increasing initial positions and that they are all moving to the left with absolute speed values between 0 and 1. We only count events that occur because the leftmost point of the 1D bounding box changes. Note that these simplifications do not asymptotically affect the results in this paper.

A necessary condition for the jth point to cause an external event is that all its *preceding* points have smaller absolute speed values, i.e. that $s_i < s_j, \ \forall i < j$. If this is the case we call s_j a left-to-right maximum. Since we are interested in an upper bound we can neglect the initial positions of the points and need only to focus on the sequence of absolute speed values $S = (s_1, \ldots, s_n)$ and count the left-to-right maxima in this sequence.

The general concept for estimating the number of left-to-right maxima within the sequence is as follows. Let f and F denote the density function and distribution function, respectively, of the noise that is added to the initial speed values. (This means $\widetilde{s}_i = s_i + \phi_i$ where ϕ_i is chosen according to density function f.)

Let $\mathbf{Pr}[\mathrm{LTR}_j]$ denote the probability that $\widetilde{s}_j$ is a left-to-right maximum. We can write this probability as

$$\mathbf{Pr}\left[\mathrm{LTR}_j\right] = \int_{-\infty}^{\infty} \prod_{i=1}^{j-1} F(x - s_i) \cdot f(x - s_j)\, \mathrm{d}x \ . \tag{1}$$

This holds since $F(x - s_i)$ is the probability that the ith element is not greater than x after the pertubation. Since all pertubations are independently from each other, $\prod_{i=1}^{j-1} F(x - s_i)$ is the probability that *all* elements preceding $\widetilde{s}_j$ are below x. Consequently, $\prod_{i=1}^{j-1} F(x - s_i) \cdot f(x - s_j)\, \mathrm{d}x$ can be interpreted as the probablity that the jth element reaches x *and* is a left-to-right maximum. Hence, integration over x gives the probability $\mathbf{Pr}[\mathrm{LTR}_j]$.

In the following we describe how to derive a bound on the above integral. First suppose that all s_i are equal, i.e., $s_i = s$ for all i. Then $\mathbf{Pr}[\mathrm{LTR}_j] = \int_{-\infty}^{\infty} F(x-s)^{j-1} \cdot f(x-s)\, \mathrm{d}x = \int_0^1 z^{j-1}\, \mathrm{d}z = 1/j$, where we substituted $z := F(x - s)$. (Note that this result only reveals the fact that the probability for the jth element to be the largest is $1/j$.)

Now, suppose that the speed values are not equal but come from some interval $[s_{\min}, s_{\max}]$. In this case $\mathbf{Pr}[\mathrm{LTR}_j]$ can be estimated by

$$\begin{aligned}
\mathbf{Pr}\left[\mathrm{LTR}_j\right] &= \int_{-\infty}^{\infty} \prod_{i=1}^{j-1} F(x - s_i) \cdot f(x - s_j)\, \mathrm{d}x \\
&\le \int_{-\infty}^{\infty} F(x - s_{\min})^{j-1} \cdot f(x - s_{\max})\, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} F(z + \delta)^{j-1} f(z)\, \mathrm{d}z \ ,
\end{aligned}$$

where we use δ to denote $s_{\max} - s_{\min}$. Let $Z^f_{\delta,r} := \{z \in \mathbb{R} \mid f(z)/f(z+\delta) \geq r\}$ denote the subset of $\mathbb{R}$ that contains all elements z for which the ratio $f(z)/f(z+\delta)$ is larger than r. Using this notation we get

$$\begin{aligned}
\mathbf{Pr}[\mathrm{LTR}_j] &\leq \int_{\mathbb{R}\setminus Z^f_{\delta,r}} F(z+\delta)^{j-1} f(z)\,\mathrm{d}z + \int_{Z^f_{\delta,r}} F(z+\delta)^{j-1} f(z)\,\mathrm{d}z \\
&\leq \int_{\mathbb{R}\setminus Z^f_{\delta,r}} F(z+\delta)^{j-1} \frac{f(z)}{f(z+\delta)} f(z+\delta)\,\mathrm{d}z + \int_{Z^f_{\delta,r}} f(z)\,\mathrm{d}z \\
&\leq r \cdot \int_{\mathbb{R}\setminus Z^f_{\delta,r}} F(z+\delta)^{j-1} f(z+\delta)\,\mathrm{d}z + \int_{Z^f_{\delta,r}} f(z)\,\mathrm{d}z \\
&\leq r \cdot \frac{1}{j} + \int_{Z^f_{\delta,r}} f(z)\,\mathrm{d}z \quad .
\end{aligned} \tag{2}$$

Now, we can formulate the following lemma.

Lemma 1 *Let f denote the density function of the noise distribution and define for positive parameters δ and r the set $Z^f_{\delta,r} \subseteq \mathbb{R}$ as $Z^f_{\delta,r} := \{z \in \mathbb{R} \mid f(z)/f(z+\delta) \geq r\}$. Further, let Z denote the probability of the set $Z^f_{\delta,r}$ with respect to f, i.e., $Z := \int_{Z^f_{\delta,r}} f(z)\,\mathrm{d}z$. Then the number of left-to-right maxima in a sequence of n elements that are perturbed with noise distribution F is at most*

$$r \cdot \lceil 1/\delta \rceil \cdot \log n + n \cdot Z \quad .$$

Proof : We are given an input sequence S of n speed values from $(0, 1]$. Let $\mathcal{L}(S)$ denote the *expected* number of left-to-right maxima in the corresponding sequence of speed values perturbed with noise distribution f. We are interested in an upper bound on this value. The following claim shows that we only need to consider input sequences of monotonically increasing speed values. □

Claim 2 *The maximum expected number of left-to-right maxima in a sequence of n perturbed speed values is obtained for an input sequence S of initial speed values that is monotonically increasing.*

From now on we assume that S is a sequence of monotonically increasing speed values. We split S into $\lceil 1/\delta \rceil$ subsequences such that the ℓth subsequence S_ℓ, $\ell \in \{1, \ldots, \lceil 1/\delta \rceil\}$ contains all speed values between $(\ell-1)\delta$ and $\ell\delta$, i.e., $S_\ell := (s \in S : (\ell-1)\cdot\delta < s \leq \ell\cdot\delta)$. Note that each subsequence is monotonically increasing.

Let $\mathcal{L}(S_\ell)$ denote the expected number of left-to-right maxima in subsequence S_ℓ. Now we first derive a bound on each $\mathcal{L}(S_\ell)$ and then we utilize $\mathcal{L}(S) \leq \sum_\ell \mathcal{L}(S_\ell)$ to get an upper bound on $\mathcal{L}(S)$.

Fix $\ell \in \{1, \ldots, \lceil 1/\delta \rceil\}$. Let k_ℓ denote the number of elements in subsequence S_ℓ. We have

$$\mathcal{L}(S_\ell) = \sum_{j=1}^{k_\ell} \mathbf{Pr}[\mathrm{LTR}_j] \quad ,$$

where $\mathbf{Pr}[\mathrm{LTR}_j]$ is the probability that the jth element of subsequence S_ℓ is a left-to-right maximum within this subsequence. We can utilize Inequality 2 for $\mathbf{Pr}[\mathrm{LTR}_j]$ because the initial speed values in a subsequence differ at most by δ. This gives

$$\mathcal{L}(S_\ell) \leq \sum_{j=1}^{k_\ell} (r \cdot \frac{1}{j} + Z) \leq r \cdot \log n + k_\ell \cdot Z \quad .$$

Hence, $\mathcal{L}(S) \leq \sum_\ell \mathcal{L}(S_\ell) \leq r \cdot \lceil 1/\delta \rceil \cdot \log n + n \cdot Z$, as desired.

2.1 Normally distributed noise

In this section we show how to apply the above lemma to the case of normally distributed noise. We prove the following theorem.

Theorem 1 *The expected number of left-to-right maxima in a sequence of n speed values perturbed by random noise from the standard normal distribution $N(0,\sigma)$ is $O(\frac{1}{\sigma}\cdot(\log n)^{3/2}+\log n)$.*

Proof : Let $\varphi(z) := \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{z^2}{2\sigma^2}}$ denote the standard normal density function with expectation 0 and variance σ^2. In order to utilize lemma 1 we choose $\delta := \frac{\sigma}{\sqrt{\log n}}$. For $z \le 2\sigma\sqrt{\log n}$ it holds that

$$\varphi(z)/\varphi(z+\delta) = e^{(\delta/\sigma^2)\cdot z+\delta^2/(2\sigma^2)} = e^{z/(\sigma\sqrt{\log n})+1/(2\log n)} \le e^3 \ .$$

Therefore, if we choose $r := e^3$ we have $Z^{\varphi}_{\delta,r} \subset [2\sigma\sqrt{\log n},\infty)$. Now, we derive a bound on $\int_{Z^{\varphi}_{\delta,r}} \varphi(z)\,\mathrm{d}z$. It is well known from probability theory that for the normal density function with expectation 0 and variance σ^2 it holds that $\int_{k\sigma}^{\infty}\varphi(z)\,\mathrm{d}z \le e^{-k^2/4}$. Hence,

$$\int_{Z^{\varphi}_{\delta,r}} \varphi(z)\,\mathrm{d}z \le \int_{2\sigma\sqrt{\log n}}^{\infty} \varphi(z)\,\mathrm{d}z \le \frac{1}{n} \ .$$

Altogether we can apply Lemma 1 with $\delta = \sigma/\sqrt{\log n}$, $r = e^3$ and $Z = 1/n$. This gives that the number of left-to-right maxima is at most $O(\frac{1}{\sigma}\cdot\log(n)^{3/2}+\log(n))$, as desired. □

2.2 Monotonic noise distributions

In this section we investigate upper bounds for general noise distributions. We call a noise distribution *monotonic* if the corresponding density function is monotonically increasing on $\mathbb{R}_{\le 0}$ and monotonically decreasing on $\mathbb{R}_{\ge 0}$. The following theorem gives an upper bound on the number of left-to-right maxima for arbitrary monotonic noise distributions.

Theorem 2 *The expected number of left-to-right maxima in a sequence of n speed values perturbed by random noise from a monotonic noise distribution is $O(\sqrt{n\log n\cdot f(0)}+\log n)$.*

Proof : Let f denote the density function of the noise distribution and let $f(0)$ denote the maximum of f. We choose $r := 2$ whereas δ will be chosen later. In order to apply Lemma 1 we only need to derive a bound on $\int_{Z^f_{\delta,r}} f(z)\,\mathrm{d}z$. Therefore, we first define sets Z_i, $i \in \mathbb{N}$ such that $\cup_i Z_i \supseteq Z^f_{\delta,r}$ and then we show how to estimate $\int_{\cup_i Z_i} f(z)\,\mathrm{d}z$.

First note that for $z+\delta < 0$ we have $f(z) < f(z+\delta)$ because of the monotonicity of f. Hence $Z^f_{\delta,r} \subseteq [-\delta,\infty)$. We partition $[-\delta,\infty)$ into intervals of the form $[(\ell-1)\cdot\delta, \ell\cdot\delta]$ for $\ell \in \mathbb{N}_0$. Now, we define Z_i to be the ith interval that has a non-empty intersection with $Z^f_{\delta,r}$. (If less than i intervals have a non-empty intersection then Z_i is the empty set.) By this definition we have $\cup_i Z_i \supseteq Z^f_{\delta,r}$ as desired.

We can derive a bound on $\int_{\cup_i Z_i} f(z)\,\mathrm{d}z$ as follows. Suppoe that all $Z_i \subset \mathbb{R}_{\ge 0}$. Let $\hat{z}_i$ denote the start of interval Z_i. Then $\int_{Z_i} f(z)\,\mathrm{d}z \le \delta\cdot f(\hat{z}_i)$ because Z_i is an interval of length δ and the maximum density within this interval is $f(\hat{z}_i)$. Furthermore it holds that $f(\hat{z}_{i+2}) \le \frac{1}{2}f(\hat{z}_i)$ for every $i \in \mathbb{N}$. To see this consider some $z_i \in Z_i \cap Z^f_{\delta,r}$. We have $f(\hat{z}_i) \ge f(z_i) > 2\cdot f(z_i+\delta) \ge 2\cdot f(\hat{z}_{i+2})$,

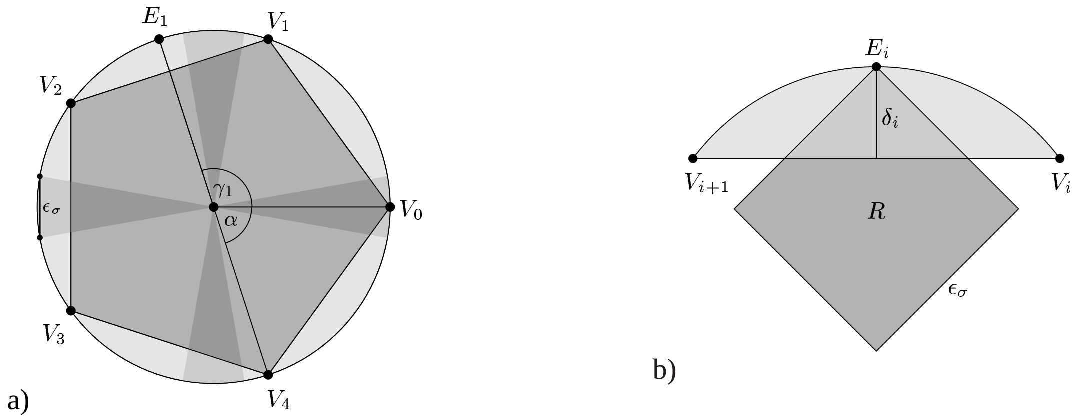

Figure 1: (a) The partitioning of the plane into different regions. If the extreme point E_i of a boundary region i falls into the shaded area the corresponding boundary region is not valid. (b) The situation where the intersection between a boundary region i and the corresponding range square R_i is minimal.

where we utilized that $z_i \in Z^f_{\delta,r}$ and that $z_i + \delta \leq \hat{z}_{i+2}$. If $Z_1 = [-\delta, 0]$ we have $\int_{Z_1} f(z)\,\mathrm{d}z \leq \delta \cdot f(0)$ for similar reasons. Now we can estimate $\int_{\cup_i Z_i} f(z)\,\mathrm{d}z$ by

$$\begin{aligned}\int_{\cup_i Z_i} f(z)\,\mathrm{d}z &\leq \sum_{i\in\mathbb{N}} \int_{Z_{2i-1}} f(z)\,\mathrm{d}z + \sum_{i\in\mathbb{N}} \int_{Z_{2i}} f(z)\,\mathrm{d}z + \int_{[-\delta,0]} f(z)\,\mathrm{d}z \\ &\leq \sum_{i\in\mathbb{N}} \frac{1}{2^{i-1}}\delta \cdot f(\hat{z}_1) + \sum_{i\in\mathbb{N}} \frac{1}{2^{i-1}}\delta \cdot f(\hat{z}_2) + \delta \cdot f(0) \\ &\leq 2\delta f(\hat{z}_1) + 2\delta f(\hat{z}_2) + \delta \cdot f(0) \leq 5\delta \cdot f(0) \quad .\end{aligned}$$

Lemma 1 yields that the number of left-to-right maxima is at most $2 \cdot \lceil \frac{1}{\delta} \rceil \cdot \log n + n \cdot 5\delta \cdot f(0)$. Now, choosing $\delta := \sqrt{\log n/(f(0) \cdot n)}$ gives the theorem.

□

3 Lower Bounds

For showing lower bounds we consider the 1D problem and map each point with initial position p_i and speed s_i to a point $P_i = (p_i, s_i)$ in 2D. We utilize that the number of external events when maintaining the bounding box in 1D is strongly related to the number of vertices of the convex hull of the P_i's. If we can arrange the points in the 2D plane such that after perturbation L points lie on the convex hull on expectation, we can deduce a lower bound of $L/2$ on the number of external events.

By this method the results of [11] directly imply a lower bound of $\Omega(\sqrt{\log n})$ for the case of normally distributed noise. For the case of monotonic noise distributions we show that the number of vertices on the convex hull is significantly larger than for the case of normally distributed noise.

We choose the uniform distribution with expectation 0 and variance σ^2. The density function f of this distribution is

$$f(x) = \begin{cases} 1/\epsilon_\sigma & |x| \leq \epsilon_\sigma/2 \\ 0 & \text{else} \end{cases} \quad , \text{ where } \epsilon_\sigma = \sqrt{12}\sigma.$$

We construct an input of n points that has a large expected number of vertices on the convex hull after perturbation. For this we partition the plane into different regions. We inscribe an ℓ-sided regular polygon into a unit circle centered at the origin. The interior of the polygon belongs to the *inner region* while everything outside the unit circle belongs to the *outer region*. Let $V_0, \ldots, V_{\ell-1}$ denote the vertices of the polygon. The ith *boundary region* is the segment of the unit circle defined by the chord V_iV_{i+1} where the indices are modulo ℓ, c.f. Figure 1a). An important property of these regions is expressed in the following observation.

Observation 3 *If no point lies in the outer region then every non-empty boundary region contains at least one point that is a vertex of the convex hull.*

In the following, we select the initial positions of the input points such that it is guaranteed that after the perturbation the outer region is empty and the expected number of non-empty boundary regions is large.

We need the following notations and definitions. For an input point j we define the range square R to be the axis-parallel square with side length ϵ_σ centered at position (p_j, s_j). Note that for the uniform distribution with standard deviation σ the perturbed position of j will lie in R. Further, the intersection between the circle boundary and the perpendicular bisector of the chord V_iV_{i+1} is called the *extremal point* of boundary region i and is denoted with E_i. The line segment from the midpoint of the chord to E_i is denoted with δ_i, c.f. Figure 1b).

The general outline for the proof is as follows. We try for a boundary region i to place a bunch of $\frac{n}{\ell}$ input points in the plane such that a vertex of their common range square R lies in the extremal point E_i of the boundary region. Furthermore we require that no point of R lies in the outer region. If this is possible it can be shown that the range square and the boundary region have a large intersection. Therefore it will be likely that one of the $\frac{n}{\ell}$ input points corresponding to the square lies in the boundary region after perturbation. Then, we can derive a bound on the number of vertices in the convex hull by exploiting Observation 3, because we can guarantee that no perturbed point lies in the outer region.

Now, we formalize this proof. We call a boundary region i *valid* if we can place input points in the described way, i.e., such that their range square R_i is contained in the unit circle and a vertex of it lies in E_i. Then R_i is called the *range square corresponding to boundary region i*.

Lemma 4 *If $\sigma \le 1/8$ and $\ell \ge 23$ then there are at least $\ell/2$ valid boundary regions.*

Proof : If $\sigma \le 1/8$ then the relationship between ϵ_σ and σ gives $\epsilon_\sigma = 2\sqrt{3}\,\sigma \le 1/2$. Let γ_i denote the angle of vector E_i with respect to the positive x-axis. A boundary region is valid iff $\sin(\gamma_i) \ge \epsilon_\sigma/2$ and $\cos(\gamma_i) \ge \epsilon_\sigma/2$. The invalid regions are depicted in Figure 1a). If $\epsilon_\sigma \le 1/2$ these regions are small. To see this let β denote the central angle of each region. Then $2\sin(\beta/2) = \epsilon_\sigma \le 1/2$ and $\beta \le 2 \cdot \arcsin(1/4) \le 0.51$. At most $\frac{\beta}{2\pi/\ell} + 1$ boundary regions can have their extreme point in a single invalid region.

Hence the total number of invalid boundary regions is at most $4(\frac{\beta}{2\pi/\ell} + 1) \le \ell/2$. □

The next lemma shows that a valid boundary region has a large intersection with the corresponding range square.

Lemma 5 *Let R_i denote the range square corresponding to boundary region i. Then the area of the intersection between R_i and the ith boundary region is at least $\min\{(\frac{4}{\ell})^4, \epsilon_\sigma^2/2\}$ if $\ell \ge 4$.*

Proof : Let α denote the central angle of the polygon. Then $\alpha = \frac{2\pi}{\ell}$ and $\delta_i = 1 - \cos(\frac{\alpha}{2})$. By utilizing the inequality $\cos(\phi) \leq 1 - \frac{1}{2}\phi^2 + \frac{1}{24}\phi^4$ we get $\delta_i \geq \frac{11}{96}\alpha^2$ for $\alpha \leq 2$. Plugging in the value for α this gives $\delta_i \geq (\frac{4}{\ell})^2$ for $\ell \geq 4$.

The intersection between the range square and the boundary region is minimal when one diagonal of the square is parallel to δ_i, c.f. Figure 1b). Therefore, the area of the intersection is at least $\delta_i^2 \geq (\frac{4}{\ell})^4$ if $\delta_i \leq \sqrt{2}\epsilon_\sigma$ and at least $\epsilon_\sigma^2/2$ if $\delta_i \geq \sqrt{2}\epsilon_\sigma$. □

Lemma 6 *If $\ell \leq \min\{\sqrt[5]{n/\epsilon_\sigma^2}, n/2\}$ then every valid boundary region is non-empty with probability at least $1 - 1/e$, after perturbation.*

Proof : We place $\frac{n}{\ell}$ input points on the center of a valid range square. The probability that none of these points lies in the boundary region after perturbation is

$$\mathbf{Pr}[\text{boundary region is empty}] \leq \left(1 - \frac{\min\{\delta_i^2, \epsilon_\sigma^2/2\}}{\epsilon_\sigma^2}\right)^{\frac{n}{\ell}},$$

because the area of the intersection is at least $\min\{\delta_i^2, \epsilon_\sigma^2/2\}$ and the whole area of the range square is ϵ_σ^2. If $\delta_i^2 = \min\{\delta_i^2, \epsilon_\sigma^2/2\}$ the result follows since

$$\frac{\epsilon_\sigma^2}{\min\{\delta_i^2, \epsilon_\sigma^2/2\}} \leq \frac{\epsilon_\sigma^2}{\delta_i^2} \leq \epsilon_\sigma^2 \cdot \ell^4 = \epsilon_\sigma^2 \cdot \ell^5/\ell \leq n/\ell \ .$$

Here we utilized that $\delta_i^2 \geq 1/\ell^4$ which follows from the proof of Lemma 5. In the case that $\epsilon_\sigma^2/2 = \min\{\delta_i^2, \epsilon_\sigma^2/2\}$ the result follows since $\frac{n}{\ell} \geq 2$. □

Theorem 3 *If $\sigma \leq 1/8$ the smoothed worst case number of vertices on the convex hull is $\Omega(\min\{\sqrt[5]{n}/\sigma, n\})$.*

Proof : By combining Lemmas 4 and 6 with Observation 3 the theorem follows immediatly if we choose $\ell = \Theta(\min\{\sqrt[5]{n}/\epsilon_\sigma, n\})$. □

4 Conclusions

We introduced smoothed motion complexity as a measure for the complexity of maintaining combinatorial structures of moving data. We showed that for the problem of maintaining the bounding box of a set of points the smoothed motion complexity differs significantly from the worst case motion complexity which makes it unlikely that the worst case is attained in typical applications.

A remarkable property of our results is that they heavily depend on the probability distribution of the random noise. In particular, our upper and lower bounds show that there is an exponential gap in the number of external events between the cases of uniformly and normally distributed noise. Therefore we have identified an important sub-task when applying smoothed analysis. It is mandatory to precisely analyze the exact distribution of the random noise for a given problem since the results may vary drastically for different distributions.

References

[1] AGARWAL, P., AND HAR-PELED, S. Maintaining approximate extent measures of moving points. In *Proceedings of the 12th ACM-SIAM Symposium on Discrete Algorithms (SODA)* (2001), pp. 148–157.

[2] BANDERIER, C., MEHLHORN, K., AND BEIER, R. Smoothed analysis of three combinatorial problems. manuscript, 2002.

[3] BAREQUET, G., AND HAR-PELED, S. Efficiently approximating the minimum-volume bounding box of a point set in three dimensions. In *Proceedings of the 10th ACM-SIAM Symposium on Discrete Algorithms (SODA)* (1999), pp. 82–91.

[4] BASCH, J., GUIBAS, L. J., AND HERSHBERGER, J. Data structures for mobile data. *Journal of Algorithms 31*, 1 (1999), 1–28.

[5] BASCH, J., GUIBAS, L. J., AND ZHANG, L. Proximity problems on moving points. In *Proceedings of the 13th Annual ACM Symposium on Computational Geometry* (1997), pp. 344–351.

[6] BLUM, A., AND DUNAGAN, J. Smoothed analysis of the perceptron algorithm. In *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms (SODA)* (2002), pp. 905–914.

[7] GUIBAS, L. J., HERSHBERGER, J., SURI, S., AND ZHANG, L. Kinetic connectivity for unit disks. *Discrete & Computational Geometry 25*, 4 (2001), 591–610.

[8] HAR-PELED, S. Clustering motion. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)* (2001), pp. 84–93.

[9] HERSHBERGER, J., AND SURI, S. Simplified kinetic connectivity for rectangles and hypercubes. In *Proceedings of the 12th ACM-SIAM Symposium on Discrete Algorithms (SODA)* (2001), pp. 158–167.

[10] KIRKPATRICK, D., SNOEYINK, J., AND SPECKMANN, B. Kinetic collision detection for simple polygons. *International Journal of Computational Geometry and Applications 12*, 1-2 (2002), 3–27.

[11] RÉNYI, A., AND SULANKE, R. Über die konvexe Hülle von n zufällig gewählten Punkten. *Zentralblatt für Wahrscheinlichkeitstheorie und verwandte Gebiete 2* (1963), 75–84.

[12] SANKAR, A., SPIELMAN, D., AND S.TENG. Smoothed analysis of the condition numbers and growth factors of matrices. manuscript, 2002.

[13] SPIELMAN, D., AND TENG, S. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. In *Proceedings of the 33rd ACM Symposium on Theory of Computing (STOC)* (2001), pp. 296–305.

[14] SPIELMAN, D., AND TENG, S. Smoothed analysis of property testing. manuscript, 2002.

[15] ZHANG, L., DEVARAJAN, H., BASCH, J., AND INDYK, P. Probabilistic analysis for combinatorial functions of moving points. In *Proceedings of the 13th Annual ACM Symposium on Computational Geometry* (1997), pp. 442–444.

Appendix

A.1 Proof of Claim 2

Proof : Consider two sequences of initial speed values, $S_1 = (s_1, \ldots, s_k, s_{k-1})$ and $S_2 = (s_1, \ldots, s_{k-1}, s_k)$ where $s_1 \le s_2 \le \cdots \le s_{k-1} < s_k$, i.e. $\gamma := s_k - s_{k-1} > 0$. It holds that

$$\Delta := \mathcal{L}(S_2) - \mathcal{L}(S_1) > 0 \ ,$$

where $\mathcal{L}(S_i)$ denotes the expected number of left-to-right maxima in the respective perturbed sequence.

To see this it suffices to consider for both sequences the probability that s_{k-1} and the probability that s_k becomes a left-to-right maximum in the perturbed sequences, because the probability for the other elements does not change. Hence,

$$\Delta = \left(\mathbf{Pr}\left[\mathrm{LTR}^{S_2}_{k-1}\right] + \mathbf{Pr}\left[\mathrm{LTR}^{S_2}_{k}\right]\right) - \left(\mathbf{Pr}\left[\mathrm{LTR}^{S_1}_{k}\right] + \mathbf{Pr}\left[\mathrm{LTR}^{S_1}_{k-1}\right]\right) \ ,$$

where LTR^S_j is the event that the j-th element in sequence S is a left-to-right maximum after perturbation. Then f and F denote the density function and distribution function, respectively, of the noise that is added to the initial speed values. Then $\mathcal{F}(z) = \prod_{i=1}^{k-2} F(z - s_i)$ is the probability that the first $k-2$ speed values do not exceed height z. Then we can write

$$\begin{aligned} \Delta \ &= \ \left(\int_{-\infty}^{\infty} \mathcal{F}(z) \cdot f(z - s_{k-1})\,\mathrm{d}z + \int_{-\infty}^{\infty} \mathcal{F}(z) \cdot F(z - s_{k-1}) \cdot f(z - s_k)\,\mathrm{d}z\right) \\ &- \ \left(\int_{-\infty}^{\infty} \mathcal{F}(z) \cdot f(z - s_k)\,\mathrm{d}z + \int_{-\infty}^{\infty} \mathcal{F}(z) \cdot F(z - s_k) \cdot f(z - s_{k-1})\,\mathrm{d}z\right) \ , \end{aligned}$$

and a linear substitution yields

$$\begin{aligned} \Delta \ &= \ \int_{-\infty}^{\infty} \mathcal{F}(z + s_k) \cdot [f(z) + F(z) \cdot f(z - \gamma)]\,\mathrm{d}z \\ &- \ \int_{-\infty}^{\infty} \mathcal{F}(z + s_k) \cdot [f(z - \gamma) + F(z - \gamma) \cdot f(z)]\,\mathrm{d}z \\ &= \ \int_{-\infty}^{\infty} \mathcal{F}(z + s_k) \cdot [(1 - F(z)) \cdot f(z) - (1 - F(z - \gamma)) \cdot f(z - \gamma)]\,\mathrm{d}z \\ &\ge \ \int_{-\infty}^{\infty} \underbrace{\mathcal{F}(z + s_k) \cdot [1 - F(z - \gamma)]}_{=:\ \hat{\mathcal{F}}(z)} \cdot (f(z) - f(z - \gamma))\,\mathrm{d}z \ . \end{aligned}$$

The last inequality holds since $\mathcal{F}(z)$ is a positive and monotonically increasing function and $F(z) > F(z - \gamma)$ for all z. Note that $\hat{\mathcal{F}}(z)$ is monotonically decreasing. And by another linear substitution we obtain

$$\begin{aligned} \Delta \ &\ge \ \int_{-\infty}^{\infty} \hat{\mathcal{F}}(z) \cdot f(z)\,\mathrm{d}z - \int_{-\infty}^{\infty} \hat{\mathcal{F}}(z) \cdot f(z - \gamma)\,\mathrm{d}z \\ &= \ \int_{-\infty}^{\infty} \hat{\mathcal{F}}(z) \cdot f(z)\,\mathrm{d}z - \int_{-\infty}^{\infty} \hat{\mathcal{F}}(z + \gamma) \cdot f(z)\,\mathrm{d}z > 0 \end{aligned}$$

which proves the lemma. □