Probabilistic Analysis for Combinatorial Functions of Moving Points

Julien Basch Harish Devarajan Piotr Indyk
Li Zhang
Computer Science Department, Stanford University, Stanford, CA94305

Revised January 26, 1998

Abstract

We initiate a probabilistic study of configuration functions of moving points. In our probabilistic model, a *particle* is given an initial position and a velocity drawn independently at random from the same distribution $\mathcal{D}$. We show that if n particles are drawn independently at random from the uniform distribution on the square, their convex hull undergoes $\Theta(\log^2 n)$ combinatorial changes in expectation, their Voronoi diagram undergoes $\Theta(n^{3/2})$ combinatorial changes, and their closest pair undergoes $\Theta(n)$ combinatorial changes. A probabilistic analysis of kinetic data structures is initiated.

1 Introduction

Given a set of n points, what is the description complexity of their convex hull? In the world of analysis of algorithms, this question is understood with an implicit "in the worst case", and the answer is $n^{\lfloor d/2 \rfloor}$ where d is the dimension of the underlying space. This is not entirely satisfactory, as this description complexity can vary tremendously depending on the positions of the points. Another type of answer is to look at the expected description complexity when the points are drawn from a given distribution. This type of analysis, initiated by Rényi and Sulanke [RS63] is valuable because this expectation is in general much smaller than the worst case, and, more importantly, because it often allows one to design algorithms that have expected running times against which worst case aware algorithms cannot compete [BCL90, Dwy91b].

In the past years, a number of papers have considered a setting where points are allowed to move along low degree algebraic trajectories [Ata85, GMR91, DG93, KTI95]. Different questions have been asked in this context. In particular, Atallah, in his pioneering work that introduced Davenport-Schinzel sequences in computational geometry [Ata85], studied the number of times the combinatorial description of the convex hull or closest pair can change, in the worst case ("dynamic computational geometry"). More recently, Basch, Guibas, and Hershberger [BGH97] have designed data structures (called *kinetic*) to maintain these attributes in an *online* setting, measuring the quality of a kinetic data structure by the ratio of the worst case number of changes to the configuration of interest to the worst case number of changes to the data structure itself, for low degree algebraic motions. This measure is not ideal and would gain to be replaced by one akin to the competitive ratio, but there is no result in this direction yet. In the meanwhile, an experimental study has been undertaken to assess the quality of these data structures in practice [BGSZ97], showing that the worst case analysis can hide vastly different results in terms of expectation when the point positions and speeds are drawn at random from some distributions. It is this study, with its surprising experimental results, that motivated the present paper.

Given these considerations, it is natural to study theoretical bounds for the expected number of changes of combinatorial functions of moving points drawn from prescribed distributions, as well as expected time bounds for kinetic data structures that maintain these combinatorial functions. For instance, it is natural to ask what is the expected number of topological changes to the Voronoi diagram under an appropriate probabilistic model.

In this paper, we obtain tight bounds on the expected number of changes to the convex hull, the Voronoi diagram, and the closest pair of points in the plane for a distribution uniform on a unit square. We also analyze the expected running time of one of the algorithms proposed in [BGH97] for the maintenance of the convex hull.

In the remaining of this section, we review previous work (probabilistic static case and worst case dynamic case), introduce our probabilistic model for moving points, and go through some preliminaries. We then devote one section for each of the combinatorial functions we studied: closest pair in Section 2, Voronoi diagram in Section 3, convex hull in section 4. Section 5 lists related results, and the last section does what you would expect.

1.1 Previous work

The average case analysis for geometric structures was initiated by Rényi and Sulanke [RS63], who looked at the expected combinatorial description, expected perimeter and area, of the convex hull of a set of points drawn independently from a uniform distribution in the unit disk. The estimation of the volume is useful for approximating convex bodies by random sampling, while the estimation of the combinatorial description has relevance for algorithm design. Surprisingly, the expected combinatorial size of the convex hull varies wildly upon the choice of distribution. Some papers have addressed these questions for spherically symmetric distributions [Efr67, Ray70, Dwy91a], some for a uniform distribution in a convex polytope [AW91, Dwy88]. A good summary can be found in [Dwy90] to complement this partial list of references. As we are going to be concerned with the unit square, let us cite the most basic result:

Theorem 1. *Consider n points drawn uniformly independently at random from a d-dimensional hypercube. Then the expected number of faces to their convex hull is* $\Theta(\log^{d-1} n)$ *.*

The Voronoi diagram of n points in the plane has linear size, but the question of the expected size becomes pertinent in higher dimensions, as, in dimension d, the Voronoi diagram has size $\Theta(n^{\lceil d/2 \rceil})$ in the worst case []. In the probabilistic setting, Dwyer shows the following (and computes the hidden constant):

Theorem 2. **[Dwy91b]** *Consider n points drawn uniformly independently at random from the d-dimensional unit ball. The complexity of the Voronoi diagram is linear in expectation.*

For the closest pair in the static case, the description complexity is of no interest, but Efron [Efr67] computed the distribution of the closest distance:

Theorem 3. **[Efr67]** *Let S be a set of n independent, identically distributed random variables in d-dimensional Euclidian space, whose common distribution is given by a bounded density function f with respect to the Lebesgue measure. Let $M_S = \min\{\|x-y\| \mid x, y \in S, x \neq y\}$. Then:*

$$\lim_{n \to \infty} (M_S > r) = \exp[-c_f n^2 r^d]$$

where

$$c_f = \frac{\pi^{d/2}}{2?\,(\frac{1}{2}d + 1)} \int_{R^d} f^2(x) dx$$

is maximized for the uniform distribution.

The study of the number of changes to combinatorial functions of moving points was initiated by Atallah [Ata85]. As a point cannot move without losing its identity, we prefer to use the term "particle" in this paper.

Definition 1. *A* particle p *in* $\mathcal{R}^d$ *is a pair* $(p_0, v_p) \in \mathcal{R}^d \times \mathcal{R}^d$. *Its components are called its* initial position *and* velocity. *Its* position at time t *is denoted and given by:*

$$p(t) = p_0 + t v_p$$

We denote by $\mathcal{P}$ the set of particles.

For a combinatorial function F on $\mathcal{R}^{d^n}$, and a set S of n particles, the number of changes to $F((p(t)_{p \in S})$ as t goes from t_0 to t_1 is denoted by $Events_F^{t_0,t_1}(S)$

Remark 1. *The quantity* $Events_F^{t_0,t_1}$ *remains the same upon transformation of all velocities or all initial positions by a translation and a uniform scaling.*

Most results having to do with moving points address the question of finding sharp bounds for the worst case, i.e.:

$$\max_{S \in \mathcal{P}^n} \text{Events}_F^{0,+\infty}(S)$$

Usually, particles are authorized to have algebraic motions of bounded degree, but the results are not qualitatively different from the fixed velocity case. The results for the combinatorial functions we consider in this paper are:

Theorem 4. [Ata85, AR92, AGHV97] *The worst case number of changes for the following combinatorial functions for n particles is:*

Convex hull in the plane	$\Theta(n^2)$
Convex hull in d-dim	$O(n^d), \Omega(n^{\lfloor d/2 \rfloor +1})$
Closest pair in d-dim	$\Theta(n^2)$
Voronoi Diagram in d-dim	$O(n^{d+1})$

The constructions for the lower bounds of the above are extremely specific. As it often happens in computational geometry, a worst case analysis gives only a partial picture.

1.2 Probabilistic setting

In this paper, we address the question of the expected number of changes to certain combinatorial functions under a given assumption on the distribution of the particles. In the remaining of this paper, we will always be in the plane.

Definition 2. *A g-random particle is a random variable whose initial position and velocity are identically and independently distributed with density g with respect to the Lebesgue measure on $\mathcal{R}^d$. We denote by $g(p)$ the density of the joint distribution for $p \in \mathcal{P}$, and by $g_t(x)$ the density of $p(t)$ for $x \in \mathcal{R}^2$ (all with respect to the Lebesgue measure).*

By remark 1, it is clear that only the shape of the density matter. Moreover,

Proposition 1. *Let F be a combinatorial function invariant by similarity, and let S be a tuple of independent g-random particles. Then:*

$$E\big[Events_F^{0,+\infty}(S)\big] = 2E\big[Events_F^{0,1}(S)\big]$$

We may choose another model, where a random particle is given by independently chosen initial and final positions with density g. Using another appropriate space-time transformation, the reader will easily see that the expected number of events in this model is the same as in our model. In [BGSZ97], a number of particles were independently drawn at random this way, and the number of events of several kinetic data structures, for several underlying distributions, were studied experimentally.

For the purpose of this paper, we will consider a probabilistic model where initial positions and velocities are taken uniformly from the unit square $0 \leq x, y \leq 1$. This model is of course equivalent (Remark 1) to the more reasonable model where velocities have coordinates drawn from the interval $[-1,+1]$. In the rest of this paper, g denotes the uniform density on the unit square, i.e., we have $g_t(x,y) = h_t(x)h_t(y)$ where:

$$h_t(z) = \begin{cases} \frac{z}{t} & \text{if } z \leq t \\ 1 & \text{if } t \leq z \leq 1 \\ \frac{1+t-z}{t} & \text{if } 1 \leq z \geq 1+t \end{cases}$$

For future reference, we note that $g_t(x,y) = \min(\frac{x}{t},1)\min(\frac{y}{t},1)$ for $0 \leq x,y \leq 1$. In the sequel, we will have to consider three cases where the density is qualitatively different: the center, the sides, and the corners. By Remark 1 once again, it will be enough to consider the bottom left corner, the bottom side, and the center (zones A, B, C in Figure 1) to obtain bounds valid for the whole range.

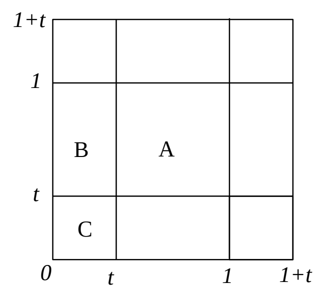

Figure 1: The distribution g_t is flat in zone A, linear in zone B, and quadratic in zone C. In general, each case may have to be considered separately.

We can consider the problem stated above in one dimension. The easiest model is the one where an initial and final positions are chosen uniformly independently at random in the segment $[0,1]$. Then the convex hull (which is the upper envelope of the trace of the particles in time-space) is a thinly disguised version of the "record breaking" problem, and the expected number of events is $\Theta(\log n)$. The closest pair and the Voronoi diagram both change when two particles meet, and there are $\Theta(n^2)$ such events in expectation.

1.3 Preliminaries

In the rest of this paper, we denote by $[P]$ the indicator of a statement P (Iversonian notation).

We briefly recall the approach of Rényi and Sulanke [RS63] for computing the expected size of the static convex hull. The idea is to consider a given pair of random points (P,Q), and to compute the probability that this pair forms an edge of the convex hull, i.e. that all other points $R_1, \cdots, R_n$, whose positions are chosen independently, are on the same side of (say above) the line passing through (P, Q).

$$\begin{aligned}
\Pr[PQ \text{ on lower convex hull}] &= \Pr[\forall i, R_i \text{ above } PQ] \\
&= E[\Pr[\forall i, R_i \text{ above } PQ | P, Q]] \\
&= E\left[\prod_i \Pr[R_i \text{ above } PQ | P, Q]\right] \\
&= E\left[(1 - A(P,Q))^{n-2}\right]
\end{aligned}$$

Where $A(P, Q)$ is the area to the left of the line passing by P, Q, which is defined almost surely. The crucial step is from line 2 to line 3: the events $[R_i$ above $PQ]$ are not independent, but they are independent conditionally on P, Q.

To compute the distribution of $A(P, Q)$, one introduces a map Φ that gives the coordinates of P and Q from the parameters of the line ℓ they define and their abscissae s_P, s_Q on this line, the change of variable formula cited later in this section (Theorem 5) gives:

$$\Pr[PQ \text{ on lower convex hull}] = \int (1 - A(\ell))^n |J\Phi(\ell, s_P, s_Q)| d\ell ds_P ds_q$$

where $A(\ell)$ is the area of the square to the left of ℓ, $J\Phi$ is the Jacobian of Φ, and the integration is taken over the space of parameters such that P and Q fall in the unit square. It is then necessary to carefully bound the Jacobian to finish the computation.

As we are going to use this a few times, let us recall the the change of variable formula. (See e.g. Billingsley [Bil86], Theorem 17.2.):

Theorem 5. *Let $\Phi : V \mapsto \Phi V$ be a one-to-one mapping of an open set V onto an open set ΦV. Suppose that Φ has continuous partial derivatives ϕ_{ij}. Then*

$$\int_V f(\Phi x)|J(x)|dx = \int_{\Phi V} f(y)dy$$

where $J(x)$ is the Jacobian of Φ at x, i.e. the determinant of the matrix of first derivatives $(\phi_{i,j})_{i,j}$.

The following integral also appears many times:

Proposition 2. *Let*

$$W = \int_0^a x^\mu e^{-cx^\nu} dx$$

Then,

$$W = \frac{1}{\mu} ? \left(\frac{\mu+1}{\nu}\right) c^{-\frac{\mu+1}{\nu}} - \Theta\left(\frac{a^{\mu-\nu+1} e^{-ca^\nu}}{c}\right)$$

This approximation is good if c is large. Otherwise, a better approximation consists in simply ignoring the exponential:

$$W \leq \frac{a^{\mu+1}}{\mu+1}$$

Notice at last that it is not necessary to be concerned with degeneracies (e.g. four points becoming collinear at some time), as the set of degenerate situations has measure 0.

The general method for the three forthcoming sections is as follows, which we describe for the convex hull for definiteness: We first focus on three particles in order to compute the expected number of times they are involved in a change to the convex hull. We identify the condition of this event (no other particle to the right of the line defined by the three particles when they become collinear), and introduce a change of variable Φ to condition upon the line of collinearity. We bound the Jacobian of the transformation and perform routine integration.

There is a technical detail: the change of variable formula requires Φ to be one-to-one, but the parameterization will cover each triplet of particles up to twice (this is because the time at which three particles become collinear is the solution of a quadratic equation). To avoid this, we introduce an intermediary space $\mathcal{M}$ on which Φ is one-to-one.

1.4 Results

We obtain the following asymptotic results for classical problems of dynamic computational geometry. For each combinatorial function, we give the expected number of times it changes under a given distribution. We recall the worst case for reference ($\beta(n)$ is an extremely slowly growing function).

Function	Changes	Theorem	worst case
Closest pair	$\Theta(n)$	6	$\Theta(n^2)$
Voronoi Diagram	$\Theta(n^{3/2})$	7	$O(n^3\beta(n))$, $\Omega(n^2)$
Convex hull	$\Theta(\log^2 n)$	8	$\Theta(n^2)$

2 Closest pair

A change in the closest pair description happens at a time t where the distance between two particles p_1, q_1 becomes equal to the distance between two particles p_2, q_2, and there is no pair at a smaller distance at that time. Hence, we first focus on a given quadruplet $P = (p_1, q_1, p_2, q_2)$ of independent g-random particles, and an additional set S of n other independent g-random particles.

For each time t, we define:

Definition 3. *Given a quadruplet* $P = (p1, q1, p2, q2)$ *of particles, a set* S *of other particles, and* $r \in \mathcal{R}^+$*, we define:*

$$\begin{aligned} C_t(P, S, r) : &\quad \forall s \in S, p \in P, \|s(t) - p(t)\| \geq r \\ D_t(S, r) : &\quad \forall s, s' \in S, \|s(t) - s'(t)\| \geq r \end{aligned}$$

The equation $\|p_1(t) - q_1(t)\| = \|p_2(t) - q_2(t)\|$ is quadratic in t, so it can have up to two solutions. Hence, we are not going to compute the probability that P triggers a change in the closest pair, but the expected number of times it can do so, i.e.:

$$E_\mu \left[|\{t \mid \|p_1(t) - q_1(t)\| = \|p_2(t) - q_2(t)\| \wedge C_t(P, S, r_t) \wedge D_t(S, r_t)\}| \right] \tag{1}$$

where $r_t = \|p_1(t) - p_2(t)\|$. It is inconvenient to work with the above expression. A more elegant replacement to (1) can be obtained if we define:

$$\mathcal{M} = \left\{ (p_1, p_2, q_1, q_2, t) \in \mathcal{P}^4 \times [0,1] \mid \|p_1(t) - q_1(t)\| = \|p_2(t) - q_2(t)\| \right\}$$

For $\xi \in \mathcal{M}$, denote P_ξ the associated quadruplet of particle parameters, t_ξ the associated time, and r_ξ the common distance between p_i and q_i at time t. We equip $\mathcal{M}$ with the measure μ induced from $\mathcal{P}^4$, i.e.

$$\int_{\mathcal{M}} f(\xi) d\mu = \int_{\xi \in \mathcal{M}} f(\xi) g(P_\xi) dP_\xi \, .$$

Where $g(P\xi)$ is a shorthand for the product of the densities of each particle of the tuple P_ξ. This measure has finite mass as $\mathcal{M}$ is locally 16-dimensional. In this case, the expected number of changes to the closest pair due to P is:

$$\Lambda = \int_{(\xi, S) \in \mathcal{M} \times \mathcal{P}^n} [D_{t_\xi}(S, r_\xi)][C_{t_\xi}(S, P_\xi, r_\xi))] g_{t_\xi}(P_\xi) g_{t_\xi}(S) dP_\xi dS \tag{2}$$

where $[\cdots]$ is the indicator of the statement enclosed.

The computations to come are rather involved but the idea is simple. We compute the expectation conditioned on the distance r between the closest pair when the event happens. Then the quantity

$$\int_S [C_{t_\xi}(S, P_\xi, r)] g_{t_\xi}(S) dS$$

is given by Theorem 3. The positions of p_1, q_1, p_2, q_2 are not independent conditioned on r, so it is necessary to do a change of variable to account for the dependencies.

Definition 4. *We parameterize the space* $\mathcal{M}$ *by the transformation* Φ *with the following parameters as domain:*

$$(p_1(0), p_2(0), u_1, u_2, r, t, \theta_1, v_1, \theta_2, v_2)$$

where $(p_i(0), u_i)$ is the initial position and velocity vector of p_i, v_i is the velocity vector of q_i, and the initial position of q_i is given by:

$$q_i(0) = p_i(t) + r \begin{pmatrix} \cos\theta_i \\ \sin\theta_i \end{pmatrix} - tv_i$$

Note that although two points in the domain of Φ can give the same quadruplet of particle parameters, they map to different points on $\mathcal{M}$, so that Φ is one-to-one. This is the main reason to introduce the space $\mathcal{M}$.

Lemma 1. *The Jacobian of Φ is equal to:*

$$J\Phi = -r^2 \begin{vmatrix} v_1^x + u_1^x & \cos\theta_1 & \sin\theta_1 & 0 \\ v_1^y + u_1^y & \sin\theta_1 & \cos\theta_1 & 0 \\ v_2^x + u_2^x & \cos\theta_2 & 0 & \sin\theta_2 \\ v_2^y + u_2^y & \sin\theta_2 & 0 & \cos\theta_2 \end{vmatrix} = O(r^2)$$

Proof. The Jacobian is easy to obtain by differentiation. All terms in the above determinant are $O(1)$. □

Lemma 2. *Let $P = (p_i)_i$ be a quadruplet of independent g-random particles and S be a set of n more. Then, with Λ defined as in (2):*

$$\Lambda = \frac{\Theta(1)}{n^3}$$

as $n \to +\infty$.

Proof. By Theorem 3, we have, for a fixed $\xi \in \mathcal{M}$:

$$\int_{S\in\mathcal{P}^n} [D_{t(\xi)}(S, r_\xi)] g_{t_\xi}(S) dS = \exp(-c_{g_{t_\xi}} n^2 r_\xi)(1 + o(1))$$

Where $c_{g_{t_\xi}}$ is maximized for a uniform distribution, and is at least $\frac{\pi}{2}$. For the upper bound, we can drop the $[C_{t_\xi}(S, P_\xi, r_\xi)]$ term in the integral (2), and we obtain:

$$\Lambda(g) \le \int_{\xi\in\mathcal{M}} \exp(-\frac{\pi}{2} n^2 r_\xi) g_{t_\xi}(P_\xi) dP_\xi$$

We now parameterize $\mathcal{M}$ by Φ, and apply the change of variable formula (Theorem 5). We obtain:

$$\begin{aligned} \Lambda &= O(1) \int e^{-\frac{\pi}{2} n^2 r} r^2 |J\Phi| \prod_{i=1,2} \left(g(p_i(0)) g(u_i) g(q_i(0)) g(v_i) dp_i dv_i d\theta_i \right) dtdr \\ &= O(1) \int_{r,t} r^2 e^{-\frac{\pi}{2} n^2 r} dtdr \end{aligned}$$

The last integral is $\Theta(1/n^3)$ by Proposition 2.

To obtain a lower bound, we restart from equation 2, and condition D_t on C_t, which changes the distribution only sightly for small r. The rest of the computation proceeds as in the upper bound, except that we notice that the Jacobian, being not always zero, can be bounded from below by $\Omega(r^2)$ on a fixed interval of integration. □

There is another scenario that is not addressed in the above: when the closest pair p_1q_1 changes to the closest pair p_1q_2. We need another lemma for this case:

Lemma 3. *Let $P = (p, q_1, q_2)$ be a triplet of independent g-random particles and S be a set of n more. The expected number of times $\|p(t) - q_1(t)\| = \|p(t) - q_2(t)\|$ and no other pair is closer, is $\Theta(1/n^3)$.*

Proof. Omitted (similar to the above). □

Theorem 6. *Let S be a set of n independent g-random particles in the plane, and let CP be the the closest pair. Then:*

$$E[Events_{CP}^{0,1}(S)] = \Theta(n)$$

Proof. We don't need to consider degeneracies (three distances become equal at the same time, or two distances become equal but the closest pair doesn't change), as the set of degeneracies has measure 0. There are two cases:

1. The closest pair changes from p_1q_1 to p_2q_2 or vice versa. There are $\binom{n}{4}$ such quadruplets and 2 method of pairing for each quadruplet., The expected number of changes for a quadruplet is $\Theta(1/n^3)$ (Lemma 2).
2. The closest pair changes from pq_1 to pq_2 or vice versa. There are $\binom{n}{2}(n-1)$ such triplets, and each triplet gives $\Theta(1/n^3)$ changes also (Lemma 3).

From which the result follows. □

3 Voronoi diagram

The Voronoi diagram of a set of particles changes when 4 particles become cocircular [GMR91]. We consider once again 4 distinguished random particles (p_1, p_2, p_3, p_4), and compute the expected number of changes to the Voronoi diagram due to these particles, in the presence of a tuple S of n other particles. The ideas are very similar to that of the previous section, and even more similar to the original treatment of Rényi and Sulanke mentioned in Section 1.3, but many technicalities arise from the treatment of the boundaries.

Definition 5. *Define $G_t(S)$ as the probability content at time t of the disk whose boundary is the circle S, and $A_t(S)$ as the conditional probability content on the circle S, i.e.*

$$\begin{aligned} G_t(S) &= \int [x \in S] g_t(x) dx \\ A_t(S) &= \int g_t(c + r(\cos\theta, \sin\theta)) r d\theta \end{aligned}$$

if S has center c and radius r.

The cocircularity test for four particles is a degree 4 algebraic equation in t. To deal with this possible multiplicity of events triggered by the same quadruplet, we define once again:

$$\mathcal{M} = \left\{ (p_1, p_2, p_3, p_4, t) \in \mathcal{P}^4 \times [0,1] \mid (p_1(t), p_2(t), p_3(t), p_4(t)) \text{ cocircular} \right\}$$

For $\xi \in \mathcal{M}$, we denote by C_ξ the circle on which the collinearity happens, by D_ξ the disk it contains, by t_ξ the time of the event, and by P_ξ the associated quadruplet of parameters. Then, we have:

$$\begin{aligned} E(\text{Events}_{VD}^{0,1}) &= \int_{\mathcal{M} \times S} [s_i(t) \notin D_\xi] g(P_\xi) dP_\xi g(S) dS \\ &= \int_{\mathcal{M}} (1 - G_t(C(\xi)))^n g(P_\xi) dP_\xi \end{aligned}$$

The integration with respect to S first corresponds to the conditional expectation on the position of the two distinguished points in the calculations of Rényi and Sulanke. We are going to parameterize $\mathcal{M}$ by the time at which the cocircularity happens, the center and radius of the circle on which it happens, and 3 additional parameters for each particle.

Definition 6. *Let Φ be the following parameterization of $\mathcal{M}$: the domain is composed of*

$$(x, y, r, t) \times \prod_i (\theta_i, v_i^x, v_i^y)$$

where (x, y) is the center and r is the radius of the circle where the particles are cocircular at t, θ_i gives the position of $p_i(t)$ on the circle, and v_i is the speed of p_i. That is, we have:

$$p_i(0) = \begin{pmatrix} x \\ y \end{pmatrix} + r \begin{pmatrix} \cos\theta_i \\ \sin\theta_i \end{pmatrix} - t v_i \,.$$

Lemma 4. *The Jacobian of Φ is:*

$$J\Phi = r^2 \begin{vmatrix} v_1^x \cos\theta_1 + v_1^y \sin\theta_1 & r\sin\theta_1 & r\cos\theta_1 & 1 \\ v_2^x \cos\theta_2 + v_2^y \sin\theta_2 & r\sin\theta_2 & r\cos\theta_2 & 1 \\ v_3^x \cos\theta_3 + v_3^y \sin\theta_3 & r\sin\theta_3 & r\cos\theta_3 & 1 \\ v_4^x \cos\theta_4 + v_4^y \sin\theta_4 & r\sin\theta_4 & r\cos\theta_4 & 1 \end{vmatrix} \tag{3}$$

In our domain of integration, all terms in this Jacobian are $O(1)$, and so:

$$|J\Phi| \leq r^4$$

Hence, we have:

Depending on the position of the circle on which the event happens, different approximations need to be performed. For $\xi \in \mathcal{M}$, let Is_Inside(ξ) be true if the center of C_ξ is inside the square $0 \leq x, y \leq 1 + t_\xi$. The function Φ above is convenient for this asymptotically dominant case, when the circle is completely contained in S_t. We give below the full proof only of this case.

Lemma 5. *Let $P = (p_i)_i$ be a quadruplet of g-random particles and S a tuple of n others. Then:*

$$\int_{\mathcal{M}} (1 - G_{t_\xi}(C_\xi))^n [\textit{Is\_Inside}(\xi)] g(P\xi) dP_\xi = \Theta(\frac{1}{n^{5/2}})$$

Proof. We use the change of variable formula (Theorem 5) with Φ, replacing the Jacobian of Φ by its upper bound. Integrating w.r.t the initial position and velocity parameters, the above is at most:

$$\int (1 - G_t(C(x,y,r)))^n A_t^4(C(x,y,r))[\text{Is\_Inside}(\xi)] dxdydrdt$$

Where $C(x,y,r)$ is the circle of center (x,y) and radius r. We consider three cases, corresponding whether the point $z = (x + r/2, y + r/2)$ belongs to zone A,B, or C of Figure 1. The result for the other zones can be deduced by symmetry and by remark 1.

The table below gives approximations for G_t (lower bound) and A_t (upper bound) , as well as the resulting integral, for each case. Integration bounds are $[0,1]$ if not given. We omit multiplicative constants for clarity, and replace the $(1-G)^n$ by the asymptotically equivalent e^{-nG}.

$z \in \cdots$	G_t	A_t	Λ	Result
A	r^2	r	$\int e^{-nr^2} r^4 drdtdxdy$	$n^{-5/2}$
B	$r\frac{x+r}{t}$	$\frac{x+r}{t}$	$\int e^{-nr\frac{x+r}{t}} \left(r\frac{x+r}{t}\right)^4 [x<t] drdtdxdy$	$n^{-5/2}$
C	$\left(r\frac{x+r}{t}\right)^2$	$\left(r\frac{x+r}{t}\right)^2$	$\int e^{-n\left(r\frac{x+r}{t}\right)^2} \left(r\frac{x+r}{t}\right)^8 [x<t][y<t] drdtdxdy$	$n^{-14/4}$

For the lower bound, consider the subset of particles such that v_i^x and v_i^y are $\Theta(1)$. As the Jacobian (Lemma 3) is nonzero almost everywhere, it is $\Theta(r^4)$ on a parameters set of nonzero mass. The expected number of events is therefore lower-bounded by an integral similar to that of case A above. □

Lemma 6.

$$\int (1 - G_t(S(x,y,r)))^n A_t^4(S(x,y,r))[\neg \textit{Is\_Inside}(\xi)] dxdydrdt = o(\frac{1}{n^{5/2}})$$

Proof. For these cases, it is necessary to bound the Jacobian more carefully, as in Dwyer [Dwy91b], by the area of the circle of cocircularity intersected with the support of the distribution. Several cases have to be considered. Details are omitted in this paper. □

Theorem 7. *Consider a set S of n independent g-random particles in the plane. VD be the Voronoi diagram. Then:*

$$E_\mu[Events_{VD}^{0,1}] = \Theta(n^{3/2})$$

Proof. The above lemmas cover all the cases. The dominant term, multiplied by about n^4 quadruplets, gives the result. □

4 Convex hull

The techniques for the convex hull are very similar to the ones used above for the Voronoi diagram, but here, the boundaries play the crucial role. We define the space to be parameterized:

$$\mathcal{M} = \left\{(p_1, p_2, p_3, t) \in \mathcal{P}^3 \times [0,1] \mid (p_1(t), p_2(t), p_3(t)) \text{ collinear}\right\}$$

and denote by t_ξ and P_ξ the time and triplet corresponding to $\xi \in \mathcal{M}$, and by a_ξ and b_ξ the positions of intersection of the line of collinearity with the x-axis and the y-axis respectively. We denote by $G_t(a, b)$ the probability that a g-random particle is below the line parameterized by a, b at time t. The expected number of changes to the lower convex hull is:

$$\int_{\mathcal{M}} (1 - G_{t_\xi}(a_\xi, b_\xi))^n g(P_\xi) dP_\xi$$

Definition 7. *We parameterize the space $\mathcal{M}$ by the variables*

$$(a, b, t, \alpha_1, v_1, \alpha_2, v_2, \alpha_3, v_3)$$

where a, b are the intersections of the line of collinearity with the (Ox) and (Oy) axes, r_i defines the position of $p_i(t)$ on this line, and $v_i = (v_i^x, v_i^y)$ is the velocity of p_i. i.e.,

$$p_i(0) = \begin{pmatrix} \alpha_i a \\ (1-\alpha_i) b \end{pmatrix} - t v_i$$

Proposition 3. *The determinant of the Jacobian of the previous transformation is*

$$\Delta(a, b, v_i, \alpha_i) = ab \left(b \begin{vmatrix} \alpha_1 & 1 & v_1^x \\ \alpha_2 & 1 & v_2^x \\ \alpha_3 & 1 & v_3^x \end{vmatrix} + a \begin{vmatrix} \alpha_1 & 1 & v_1^y \\ \alpha_2 & 1 & v_2^y \\ \alpha_3 & 1 & v_3^y \end{vmatrix} \right)$$

If the particles are drawn from the unit square distribution, we have:

$$\Delta(a, b, v_i, \alpha_i) = O\left(ab \min(\frac{ab}{t}, a+b) \right)$$

Proof. The Jacobian is easy to obtain. For the upper bound, notice that, as $p_i^x(t) = \alpha_i a$, the x-speed is at most $\alpha_i a/t$ (and also bounded by 1), and $\alpha_i \leq 1$. Therefore, bounding a determinant by the product of the norms of the column vectors:

$$\begin{vmatrix} \alpha_1 & 1 & v_1^x \\ \alpha_2 & 1 & v_2^x \\ \alpha_3 & 1 & v_3^x \end{vmatrix} \leq 3 \max(v_1^x, v_2^x, v_3^x) \leq 3 \min(a/t, 1)$$

Hence the determinant is at most:

$$3ab\,(b \min(a/t, 1) + a \min(b/t, 1))$$

which can be put in the form announced. □

Lemma 7. *Let (p_1, p_2, p_3) be a triplet of independent g-random particle and S be a set of n others. Then:*

$$\int_{\mathcal{M}} (1 - G_{t_\xi}(a_\xi, b_\xi))^n g(P_\xi) dP_\xi = \Theta\left(\frac{\log^2 n}{n^3}\right)$$

Proof. With the change of variable Φ, we have:

$$\int_{\mathcal{M}} (1 - G_{t_\xi}(a_\xi, b_\xi))^n g(P_\xi) dP_\xi = \int (1 - G_t(a, b))^n \Delta(a, b, \alpha_i, v_i) \prod_i g(p_i) d\alpha_i dv_i da db dt$$

We consider the case $0 < a < b < t < 1$. In this case, we have the following bounds:

$$\begin{aligned} G_t(a, b) &= \Theta(\frac{a^2 b^2}{t^2}) \\ \int_0^1 g_t\begin{pmatrix} \alpha a \\ (1-\alpha) b \end{pmatrix} d\alpha &= \Theta\left(\frac{ab}{t^2}\right) \end{aligned}$$

As usual, we replace the Jacobian by the upper bound found in Proposition 3, integrate with respect to the velocities, then with respect to the α_i's. We obtain the upper bound:

$$\int (1 - \frac{a^2 b^2}{t^2}) \frac{a^2 b^2}{t} \left(\frac{ab}{t^2}\right)^3 [0 < a < b < t < 1] da db dt$$

Integrating with respect to a (Proposition 2), we obtain

$$\int \frac{b^5}{t^7} \min\left(\left(\frac{t^2}{nb^2}\right)^3, b^6\right) [0 < b < t < 1] db dt$$

which gives $\Theta(\log^2 n / n^3)$. The other cases give only lower order terms and are omitted in this paper. The lower bound argument follows that of the Voronoi diagram. □

Theorem 8. *Let CH be the combinatorial description of the convex hull of a set of points. If S is a set of n g-random particles where g is uniform on the square, then:*

$$E[Events^{0,1}_{CH}(S)] = \Theta(\log^2 n)$$

Proof. We apply the above lemma to the $\binom{n}{3}$ triplets of S, which gives a bound of $\Theta(\log^2 n)$ on the number of changes to the lower convex hull. □

5 Related results

The results we obtained for the expected number of changes to the convex hull of particles drawn from the uniform square distribution directly imply an expected linear bound for the convex hull kinetic data structure described in [BGH97]. This data structure divides the point set arbitrarily into a blue and a red half, recursively computes the value and red convex hulls, and maintains a set of certificates between edge-vertex pairs and edge-edge pairs to certify the convex hull of the whole set. If N_b and N_r are the (random) total number of edges that ever appear on the blue and red convex hull, the number of events involved at the top level is $O(N_b N_r)$. As these two quantities are independent, the expectation is $E(N_b N_r) = E(N_b)E(N_r) = \Theta(\log^4 n)$ in the case of the square distribution. The expected running time $T(n)$ of the whole data structure thus obeys the recurrence $T(n) = O(\log^4 n) + 2T(n/2)$, which solves to $T(n) = O(n)$. It is easy to see that this is also a lower bound. This result will hold for any reasonable distribution where the number of changes to the convex hull is sublinear in expectation. The results above make it clear that it is not a good idea to use the Delaunay triangulation (the dual of the Voronoi diagram) to maintain the convex hull.

The probabilistic analysis can be applied to all configuration functions that are typically looked at in the setting of moving points, and to the kinetic data structures that are used to maintain them.

For instance, as the minimum spanning tree is a subgraph of the Delaunay triangulation, the result on the Delaunay triangulation coupled with a standard batching argument imply that, in our model, the MST in the plane changes $O(n^{5/2})$ times on average. Although this is probably not a tight bound, it is to be compared with the best known worst case upper bound of $O\left(n^3 2^{\alpha(n)}\right)$ [KTI95].

The furthest pair, on the other hand, is always constituted of two points that are on the convex hull. From this and Devroye's moment inequalities [Dev83], we deduce that the number of changes to the furthest pair is polylogarithmic in expectation, for the square distribution.

6 Conclusion

In this paper we initiated the study of average case behavior of some combinatorial properties defined on points moving on random trajectories. We obtained tight bounds for convex hull, closest pair and Voronoi diagram, and for the kinetic data structure of [BGH97] to maintain the convex hull.

The computations for the closest pair can easily be extended to arbitrary dimensions, and to any distribution with bounded density. In d dimensions, the closest pair changes $\Theta(n^{\frac{2}{d}})$ times. The computations for the Voronoi diagram are more involved, due to the treatment of the boundaries. However, it is intuitively clear that the asymptotic result holds for any reasonable distribution. Hence, there must be a way to prove the result in a much simpler fashion. In d dimensions, the "A" case of Lemma 5 can be easily seen to be $\Theta(n^{1+\frac{1}{d}})$, but our method becomes too cumbersome for the boundaries.

For the above two functions, it is worth noticing that the rate of events is about constant over time, while this is not the case for the convex hull: most changes happen at the beginning.

We now review several kinetic data structures for the maintenance of the closest pair. The Delaunay triangulation can be used to maintain the closest pair, and will do it in roughly $n^{3/2}$ time in our probabilistic model. How do other methods compare? One solution is to maintain the L_1 Delaunay triangulation [CD85], as one of its edges is the closest pair, but our techniques can be used to show that it undergoes roughly the same number of changes as the L_2 Delaunay. Another method was proposed in [BGH97] and modified in [BGSZ97], whose average running time is experimentally also roughly $\Theta\left(n^{3/2}\right)$. At last, one may consider a more straightforward algorithm, which cuts the square (say) into n cells and tracks every particle as it goes from cell to cell [KSG]: although quadratic in the worst case, it is easy to see that the average case of this algorithm in our probabilistic model is also precisely $\Theta\left(n^{3/2}\right)$. There is probably a good reason for that.

In conclusion, here are the four most interesting problems that are raised by this work.

- Find a way to compute the expected number of events for the Voronoi diagram that applies to any reasonable distribution and extend it to arbitrary dimensions,
- Compute the expected number of changes to the Euclidian Minimum Spanning Tree,
- Compute the expected number of events processed by the kinetic data structure of [BGH97] for the maintenance of the closest pair,
- Prove or disprove that no kinetic data structure for the maintenance of the closest pair can process less than $\Omega(n^{3/2})$ events in expectation.

Acknowledgments

We would like to express our gratitude to Young-Eun Choi, Alexei Novikov and Suresh Venkatasubramanian for several very useful discussions, and to Leo Guibas for his constant encouragements.

References

[AGHV97] P. K. Agarwal, L. Guibas, J. Hershberger, and E. Veach. Maintaining the extent of a moving point set. In *Proceedings of the 5th Workshop on Algorithms and Data Structures*, pages 31–44. Springer-Verlag, 1997. Lecture Notes in Computer Science 1272.

[AR92] G. Albers and T. Roos. Voronoi diagrams of moving points in higher dimensional spaces. In *Proc. 3rd Scand. Workshop Algorithm Theory*, volume 621 of *Lecture Notes Comput. Sci.*, pages 399–409. Springer-Verlag, 1992.

[Ata85] M. J. Atallah. Some dynamic computational geometry problems. *Comput. Math. Appl.*, 11:1171–1181, 1985.

[AW91] F. Affentranger and J. A. Wieacker. On the convex hull of uniform random points in a simple d-polytope. *Discrete Comput. Geom.*, 6:291–305, 1991.

[BCL90] J. L. Bentley, K. L. Clarkson, and D. B. Levine. Fast linear expected-time algorithms for computing maxima and convex hulls. In *Proc. 1st ACM-SIAM Sympos. Discrete Algorithms*, pages 179–187, 1990.

[BGH97] J. Basch, L. J. Guibas, and J. Hershberger. Data structures for mobile data. In *8th Symposium on Discrete algorithms*, pages 747–756, 1997.

[BGSZ97] J. Basch, L. J. Guibas, C. D. Silverstein, and L. Zhang. A practical evaluation of kinetic data structures. In *Proc. 13th Annu. ACM Sympos. Comput. Geom.*, pages 388–390, 1997.

[Bil86] P. Billingsley. *Probability and measure.* John Wiley & Sons, 2nd edition edition, 1986.

[CD85] L. P. Chew and R. L. Drysdale, III. Voronoi diagrams based on convex distance functions. In *Proc. 1st Annu. ACM Sympos. Comput. Geom.*, pages 235–244, 1985.

[Dev83] L. P. Devroye. Moment inequalities for random variables in computational geometry. *Computing*, 30:111–119, 1983.

[DG93] O. Devillers and M. Golin. Dog bites postman: Point location in the moving Voronoi diagram and related problems. In *Proc. 1st Annu. European Sympos. Algorithms (ESA '93)*, volume 726 of *Lecture Notes in Computer Science*, pages 133–144. Springer-Verlag, 1993.

[Dwy88] R. A. Dwyer. On the convex hull of random points in a polytope. *J. Appl. Probab.*, 25(4):688–699, 1988.

[Dwy90] R. A. Dwyer. Kinder, gentler average-case analysis for convex hulls and maximal vectors. *SIGACT News*, 21(2):64–71, 1990.

[Dwy91a] R. A. Dwyer. Convex hulls of samples from spherically symmetric distributions. *Discrete Appl. Math.*, 31(2):113–132, 1991.

[Dwy91b] R. A. Dwyer. Higher-dimensional Voronoi diagrams in linear expected time. *Discrete Comput. Geom.*, 6:343–367, 1991.

[Efr67] B. Efron. The problem of the two nearest neighbors (abstract). *Ann. Math. Statist.*, 38:298, 1967.

[GMR91] L. Guibas, J. S. B. Mitchell, and T. Roos. Voronoi diagrams of moving points in the plane. In *Proc. 17th Internat. Workshop Graph-Theoret. Concepts Comput. Sci.*, volume 570 of *Lecture Notes in Computer Science*, pages 113–125. Springer-Verlag, 1991.

[KSG] D.-J. Kim, S.-Y. Shin, and L. J. Guibas. Fast collision detection among multiple moving spheres. This volume.

[KTI95] N. Katoh, T. Tokuyama, and K. Iwano. On minimum and maximum spanning trees of linearly moving points. *Discrete Comput. Geom.*, 13:161–176, 1995.

[Ray70] H. Raynaud. Sur l'enveloppe convexe des nuages de points aléatoires. *J. Appl. Prob.*, 7:35–48, 1970.

[RS63] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gerwähten Punkten I. *Z. Wahrsch. Verw. Gebiete*, 2:75–84, 1963.