The Voronoi tessellation generated from different distributions of seeds

L. Zaninetti
Istituto di Fisica Generale, Università degli Studi di Torino, Via Pietro Giuria 1, 10125 Turin, Italy

Received 11 February 1991; revised manuscript received 26 September 1991; accepted for publication 9 March 1992
Communicated by A.R. Bishop

We first present a general algorithm able to extract the properties of the Voronoi tessellation through an approximate method. We then explore how some typical parameters vary when different processes to generate the system are considered.

1. Introduction

The Voronoi tessellation in a plane or in a space has recently aroused renewed interest due to its applications in various areas ranging from astrophysics (see refs. [1,2]) to the representation of properties of materials (see ref. [3]). It turns out that some typical parameters like the area distribution in two dimensions and the volume distribution in three dimensions depend strongly on the kind of distribution adopted for the seeds. In particular we will adopt three processes (random, eigenvalues from complex matrices and Sobol) to generate the seeds (see section 2). The adopted probability distribution to fit the normalized histogram of frequencies (the gamma variate, the normal and the lognormal) are described in section 3. The scanning algorithm is presented in section 4, the numerical results are in section 5 and in section 6 we present an astrophysical application.

2. The seeds

The results of the scanning are strongly dependent on the distribution of the seeds. We have used three different processes to generate them.

2.1. The random points

The points are generated independently on the X and Y axes in 2D (adding the Z axis in 3D) through the subroutine G05CAF from the NAG library that returns a pseudo-random real number taken from a uniform distribution between 0 and 1.

2.2. The Sobol points

In order to have the seed points distributed as uniformly as possible in a two-dimensional space (or n-dimensional) we generate quasi-random numbers. The Sobol sequence generates numbers between 0 and 1 directly as binary fractions of length $w$ bits, from a set of $w$ special binary fractions, $y=1, 2, ..., w$, called direction numbers. For practical purposes we used the subroutine SOBSEQ as described in ref. [4].

2.3. The eigenvalue points

We start with a random $N \times N$ complex matrix as described in ref. [5]. The matrix elements are given by $x+iy$ where $x$ and $y$ are pseudo-random real numbers taken from a normal (Gaussian) distribution with mean 0 and standard deviation $1/\sqrt{2}$. For this last purpose we have used the subroutine G05DDF from the NAG library. Once the complex elements are obtained we diagonalize the complex matrix us-
ing the subroutine FOZAJF from the NAG library. The seed points have coordinates $x$ and $y$ corresponding to the real and imaginary parts of the complex eigenvalues.

3. Interpolating probability functions

Once a frequency distribution $F_i(x_i)$ was obtained, where $F_i$ is the number of elements comprised between $x_i - \Delta x$ and $x_i + \Delta x$, where in our case $x$ represents the area/mean area or the volume/mean volume, we fitted it through three probability distributions: the normal (Gaussian), the lognormal and the gamma variate.

3.1. The lognormal distribution function

We can consider the following normalized distribution function,

$$f_{LN}(x) = \frac{1}{\sqrt{2\pi}\ln \sigma} \exp \left( - \frac{(\ln x - \ln \bar{x})^2}{2(\ln \sigma)^2} \right).$$

The parameters $\bar{x}$ and $\sigma$, which characterize the distribution, denote the statistical median and the standard deviation as defined by

$$\ln \bar{x} = \frac{\sum n_i \ln x_i}{\sum n_i}.$$  

$$\ln \sigma = \left( \frac{\sum n_i (\ln x_i - \ln \bar{x})^2}{\sum n_i} \right)^{1/2}.$$  

where $n_i$ is the number fraction of particles in a size interval centered around $x_i$. If now we want to speak of a theoretical distribution we consider a particle density clustered around some value $x$, that is the number of elements occurring within some size interval $\Delta x_i$, expressed as a linear density $n_i/\Delta x_i$. This quantity however is still not the one represented by $f_{LN}$. In fact the physical significance of $f_{LN}$ is that it is numerically equal to $n_i x_i/\Delta x_i$, that is the quantity obtained by weighting this particle density by the mid-interval size $x$ [6]. The theoretical frequency distribution of $x$ is given by

$$n_i = N \frac{\Delta x_i}{x} f_{LN}.$$  

3.2. The gamma variate distribution

Once the normalized distribution is obtained we can fit it, following ref. [7], with the curve

$$H(z; c) = \frac{c}{\Gamma(c)} (cz)^{c-1} e^{-cz}.$$  

This fit is obtained through a subroutine for finding an unconstrained minimum of the squares of $i$ nonlinear functions in one variable ($c$).

3.3. The $\chi^2$ test

We easily compute the $\chi^2$ through the standard formula

$$\chi^2 = \sum_{i=1}^{N} \frac{(F_i - n_i)^2}{n_i},$$

and through the $\chi^2$ probability function we compute the confidence level (expressed as a percentage) connected with the probability that the sum of squares of $\nu$ (number of degrees of freedom) will be larger than $\chi^2$; this quantity is reported in the figures and tables.

4. The two-dimensional scan

We consider a lattice $(\text{pixel})^2$ and associate to every point a little area (small box) $l \times l$ where $l$=side/pixel. The maximum value of pixel depends on the computer's dynamical memory (in our case pixel could be 500 or less) and side is the length of the square. In order to find the lattice points (little square) that belong to the Voronoi cells we use the following two-step algorithm:

(i) At the center of every pixel we compute the distances from the nuclei. We then sort the obtained vectors in ascending order. Then we select the possible candidates of the lattice points giving the condition that the difference between the first two elements of the sorted array is less than $\sqrt{2} l$

(ii) Once the first class of candidates is obtained we draw a line between the two nearest nuclei. In the midpoint we draw a perpendicular line (perpendicular bisector) and we check if it intersects the little square associated with the considered pixel or not.
The code was then run under the following conditions:
(a) A boundary square, named big box, enclosing the set of nuclei;
(b) In order to simulate the boundary conditions we considered a smaller area (small box) and in every lattice point we computed the distance from the nuclei;
(c) We considered only closed polygons belonging to the nuclei of an area smaller than the small box. The exact ratios between the three areas are shown in the figures.

A typical run using the seeds as deduced from the random numbers is given in fig. 1, the frequency histogram and the relative best fits using the three distributions for the area distribution are given in fig. 2.

In order to understand better the differences between the three types of seeds we review in table 1 the family parameters. We should stress that in a previous analysis [7] a value of $c \approx 4$ has been extracted for the $\Gamma$ distribution and random seeds. We find a similar value, $c = 3.37$, but the lognormal now fits the data at the same confidence level. The Sobol and the eigenvalue points produce a more regular partition of the plane and the value of $c$ is increasing progressively. When the eigenvalue points (the more regular of the adopted seeds) are considered the normal distribution fits the data with a sufficient confidence level.

5. The three-dimensional scan

We now work on a 3D lattice $L_{k,m,n}$ of (pixel)$^3$ elements. The prescriptions are now:
(a) A big boundary volume:
(b) To take account of the boundary conditions we considered a smaller volume in which we placed our lattice;
(c) We considered only closed polyhedra belonging to the nuclei of a smaller volume than that in which the lattice was placed. The exact ratios between the three volumes are given in the figures

5.1. The 3D statistics

Here also we analyzed the distribution of the polyhedron volume and like in the previous sections we use two types of seeds: random ones and those pro-
Table 1
Family parameters family for three different seed point distributions concerning the area distribution.

<table>
<thead>
<tr>
<th>Seeds</th>
<th>Lognormal</th>
<th>Normal</th>
<th>Gamma variate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln $\bar{x}$</td>
<td>ln $\sigma$</td>
<td>CL (%)</td>
</tr>
<tr>
<td>random</td>
<td>-0.12</td>
<td>0.50</td>
<td>27.6</td>
</tr>
<tr>
<td>Sobol</td>
<td>-0.03</td>
<td>0.25</td>
<td>0.1</td>
</tr>
<tr>
<td>eigenvalues</td>
<td>-0.03</td>
<td>0.26</td>
<td>13.4</td>
</tr>
</tbody>
</table>

Table 2
Family parameters for three different seed point distributions concerning the volume distribution.

<table>
<thead>
<tr>
<th>Seeds</th>
<th>Lognormal</th>
<th>Normal</th>
<th>Gamma variate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln $\bar{x}$</td>
<td>ln $\sigma$</td>
<td>CL (%)</td>
</tr>
<tr>
<td>random</td>
<td>-0.09</td>
<td>0.43</td>
<td>1.71</td>
</tr>
<tr>
<td>sobol</td>
<td>-0.02</td>
<td>0.23</td>
<td>0.22</td>
</tr>
</tbody>
</table>

duced by the Sobol process. We do not have a three-dimensional counterpart of the complex eigenvalues from random complex matrices. For the distribution of the volumes belonging to the polyhedra, the relative fitting is shown in Fig. 3.

The experimental frequencies here also are better approximated with the gamma variate, with $c$ varying from 5.38 for a random process to $\approx 20$ for the Sobol process. The value obtained for the gamma variate for the volumes should be compared with the value 6 as deduced in ref. [7] and successively refined to 5.5 due to a change in the generator of random numbers [8]. Here also shifting toward more regular distributions, the value of $c$ increases.

6. The astrophysical megawalls

In order to test the reliability of our codes we have analyzed the puzzle of the periodic walls that characterize the galaxy distribution in the universe.

In a recent beam pencil survey along the galactic poles peaks appear when the number of pairings of galaxies at a particular distance apart is plotted against the distance between them (see ref. [9]). The period between one wall and the other one turns out to be $P = 128$ Mpc and the associated standard deviation $\approx P/10$. We can now perform the following Monte Carlo experiment: we scan the volume through many lines all passing through the center of the box, the orientation of each line being chosen with a random process. We then count how many faces are

![Fig. 3. Volume distribution using random seeds. Lognormal: ln $\bar{x} = -0.09$, ln $\sigma = 0.43$, CL = 1.7189%. Normal: $\bar{x} = 1.00$, $\sigma = 0.40$, CL = 0.2159%. Gamma variate: $c = 5.38$, CL = 19.4650%. 149 selected box nuclei, 550 big box nuclei, 200 pixels, side = 100.00, volume big box = 2.00 $\times$ volume small box = 2.00 $\times$ selected nuclei volume.](image-url)
Fig. 4. Linear scanning using Sobol seeds. $x=0.5496$, $\sigma=0.1029$, 200 experiments, maximum number of periods=22, minimum = 8, 5339 small box nuclei, 8000 big box nuclei, 400 pixels, volume big box = 1.50 × volume small box, suggested ratio = 0.20, connected probability = 0.0003.

Table 3

<table>
<thead>
<tr>
<th>sdv/period</th>
<th>Probability random</th>
<th>Probability Sobol</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>$7.74 \times 10^{-6}$</td>
<td>$6.22 \times 10^{-6}$</td>
</tr>
<tr>
<td>0.2</td>
<td>$3.20 \times 10^{-6}$</td>
<td>$3.39 \times 10^{-6}$</td>
</tr>
<tr>
<td>0.3</td>
<td>$6.11 \times 10^{-6}$</td>
<td>$7.63 \times 10^{-6}$</td>
</tr>
<tr>
<td>0.4</td>
<td>$5.50 \times 10^{-6}$</td>
<td>$7.23 \times 10^{-6}$</td>
</tr>
<tr>
<td>0.5</td>
<td>0.24</td>
<td>0.31</td>
</tr>
<tr>
<td>0.6</td>
<td>0.58</td>
<td>0.68</td>
</tr>
<tr>
<td>0.7</td>
<td>0.87</td>
<td>0.92</td>
</tr>
<tr>
<td>0.8</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>0.9</td>
<td>0.99</td>
<td>0.999</td>
</tr>
</tbody>
</table>

crossed and the distance between them.

In each direction we compute the ratio standard deviation/averaged period and we finally plot the frequencies. Fitting these frequencies through a Gaussian distribution we easily find the probability to have a ratio as low as a given number. Considering, for example, 8000 nuclei and 200 random lines the probability to have a ratio of standard deviation/period lower than 0.2 is $3.2 \times 10^{-4}$ for the random seeds and $3.4 \times 10^{-4}$ for the Sobol seeds.

This means that the probability to find a regular spacing turns out to be very low, adopting our point of view. We show in table 3 a synopsis of the expected probability versus this ratio.

7. Conclusions

Using a random distribution of seeds as a reference point we examined how other processes that are covering the plane/space in a more uniform way modify quantities like the distribution of the polygon areas or polyhedron volumes. From the inspection of the $\chi^2$ it turns out that both the gamma variate and the lognormal distributions fit with an acceptable confidence level the obtained frequencies. Conversely when more regular seeds are considered the normal distribution becomes adequate.

The developed algorithm was then applied to extract the theoretical probability connected with the regular periodicity of the "megawalls" and it turns out that this phenomenon has a low probability to be found.

References