An Alternative Method for Characterization and Comparison of Plant Root Shapes

A thesis submitted to the
College of Graduate and Postdoctoral Studies
in partial pulfillment of the requirements
for the degree of Master of Science
in the Department of School of Environment and Sustainability
University of Saskatchewan
Saskatoon

By
Yujie Pei

Contents

Existing Morphological Descriptors for Root Systems

An Alternative Mathematical Method for Shape Description

2.1 Kac's Idea: Can One Hear the Shape of a Drum? [5]

2.1.1 Interpretations of Kac's Problem

- When the drum vibrates, one can hear the sound, which is composed of tones of various frequencies. How much can shape features be inferred from hearing a discrete spectrum of pure tones produced by a drum?
- If a complete sequence of eigenvalues of the Dirichlet problem for the Laplacian can be obtained precisely, will people determine the shape of a planar?

2.1.2 Problem Statement

- Consider a simply connected membrane Ω in the Euclidean space bounded by a smooth convex curve $\partial\Omega$ (e.g. a drum without any holes)
- Find function ϕ on the closure of Ω, which vanishes at the boundary $\partial\Omega$, and a number λ satisfying $-\Delta\phi = \lambda\phi$.
  - Δ is the Laplace operator. e.g. $\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$ in Cartesian coordinate system.
  - If there exists a solution $\phi \neq 0$, the corresponding λ is defined as a Dirichlet eigenvalue.
  - For each domain Ω, there has a sequence of eigenvalues $\lambda_1, \lambda_2, \lambda_3, ...$ corresponding to a set of eigenfunction $\phi_1, \phi_2, \phi_3, ...$
  - ϕ_k form an orthonormal basis of $L^2(\Omega)$ of real valued eigenfunctions; the corresponding discrete Dirichlet eigenvalues are positive ($\lambda_k \in \mathbb{R}^+$).
- An important function [4]:

$$h(t) = \sum_{k=1}^{\infty} e^{-\lambda_k t} \tag{2.1}$$

  - It is a Dirichlet series.
  - It is called the spectral function or the heat trace.
  - It is smooth and converges for every $t > 0$.

2.1.3 Summarize the Results of Kac's Idea

$$h(t) = \sum_{k=1}^{\infty} e^{-\lambda_k t} \sim \frac{|\Omega|}{2\pi t} - \frac{L}{4}\frac{1}{\sqrt{2\pi t}} + \frac{1}{6} \tag{2.2}$$

- As $t \to 0^+$, the leading terms of the asymptotic expansion of $h(t)$ imply the geometrical attributes of Ω

  - the total area
  - the perimeter
  - the curvature

- If the domain Ω has the polygonal boundary, the third term shows in the information about the interior angles of the polygon [4].

2.1.4 Conclusion

2.1.4.1 Advantages

- Kac proposed a novel analytical mathematical method for the shape description without using measuring tools, e.g. rulers.
- Other mathematicians extended Kac's idea in exploring the geometrical information of more complex domains with various boundary conditions [6][3][2] [9][7].

2.1.4.2 Limitations

- It is only available for the convex domain, which has a smooth or piecewise smooth boundary.
- Except in very few cases (i.e. rectangular, disk, certain triangles), the complete sequence of eigenvalues λ_k can not be calculated [4].
- Only the first few terms in the asymptotic expansion of $h(t)$ are explicitly available.

LRWs in Artificial Images

	test_statistic	p
Logrank	137.23	0.0
Tarone-Ware	134.31	0.0
Gehan-Breslow	123.83	0.0
Fleming-Harrington	123.83	0.0

Table 3.1: Survival functions for circle and rectangle are statistically different since p values equal zeros.

The fixed-time step Monte Carlo simulation, LRWs, has been validated in the annulus in Appendix B by comparing the analytical and numerical survival function. The further validation of LRWs is to distinguish the geometries and explore their structural features from both short and long time behaviours of the survival function.

3.1 Circle and Rectangle

Given two simple convex shapes with the same area, circle and rectangle, we are interested in how and whether their corresponding survival curves differ from each other. For the equal-area geometries, rectangle and circle, Eq. 2.2 indicates that the survival function of the former decays faster than the latter as the time approaches zero.

The preliminary step of testing the research hypothesis is to generate two black-and-white images with the same dimensions as shown in Fig. C.1. In the binary images, circle and rectangle have an equal number of white pixels. For simplicity, the centroid of shapes located at the center of the image. Then, simulating LRWs in the images and estimating survival functions by Kaplan-Meier estimator.

3.1.1 Output Analysis

The differences between survival functions for the circle and rectangle are not visible. Moreover, the approximate 95% confidence intervals of the survival functions overlap. In this case, non-parametric statistical tests can be used to compare entire survival distributions and assess their dissimilarities. The logrank test has maximum power if the proportional hazards assumption is satisfied.

3.1.2 Conclusion

Although the proportional hazard assumption test is failed as shown in Fig. **??**, the weighted logrank tests indicate that the null hypothesis should be rejected. In conclusion, LRWs is an alternative tool to quantify and distinguish the geometries in the $2-$ dimensional image without measuring the predefined shape descriptors.

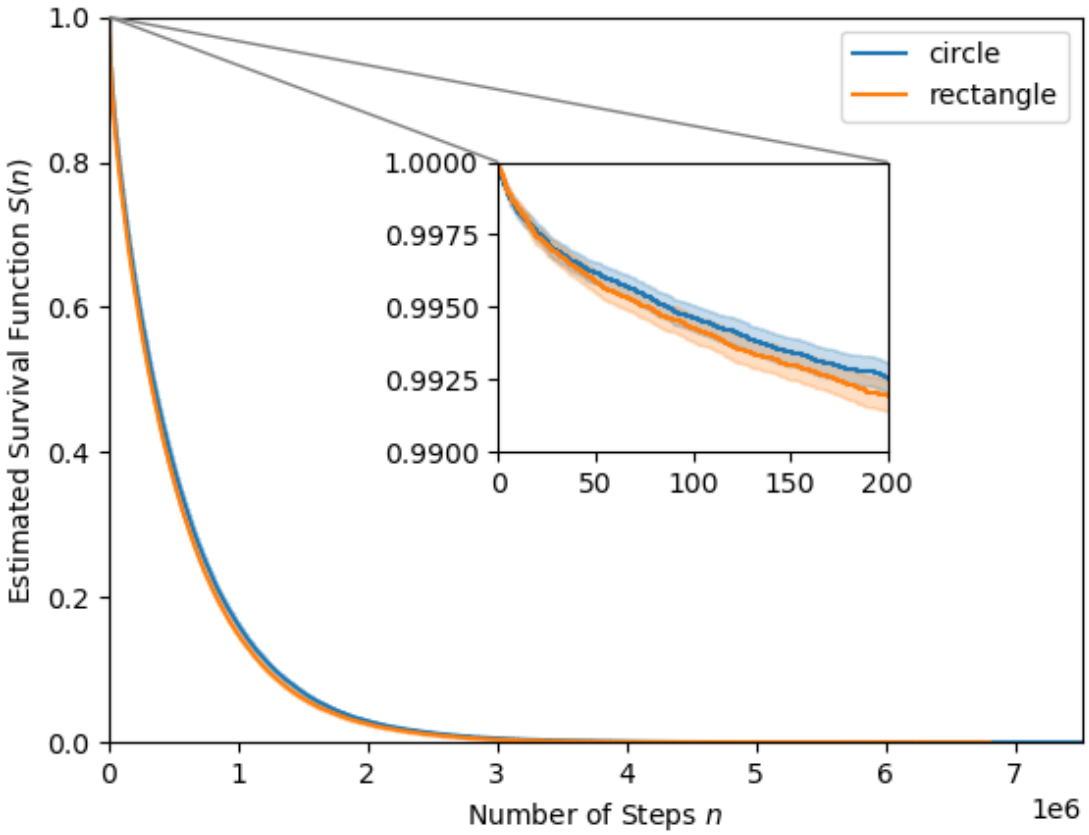


Figure 3.1: In the inset plot, the decay rate of the survival function for the rectangle is slightly larger than for the circle, which coincides with the theoretical result.

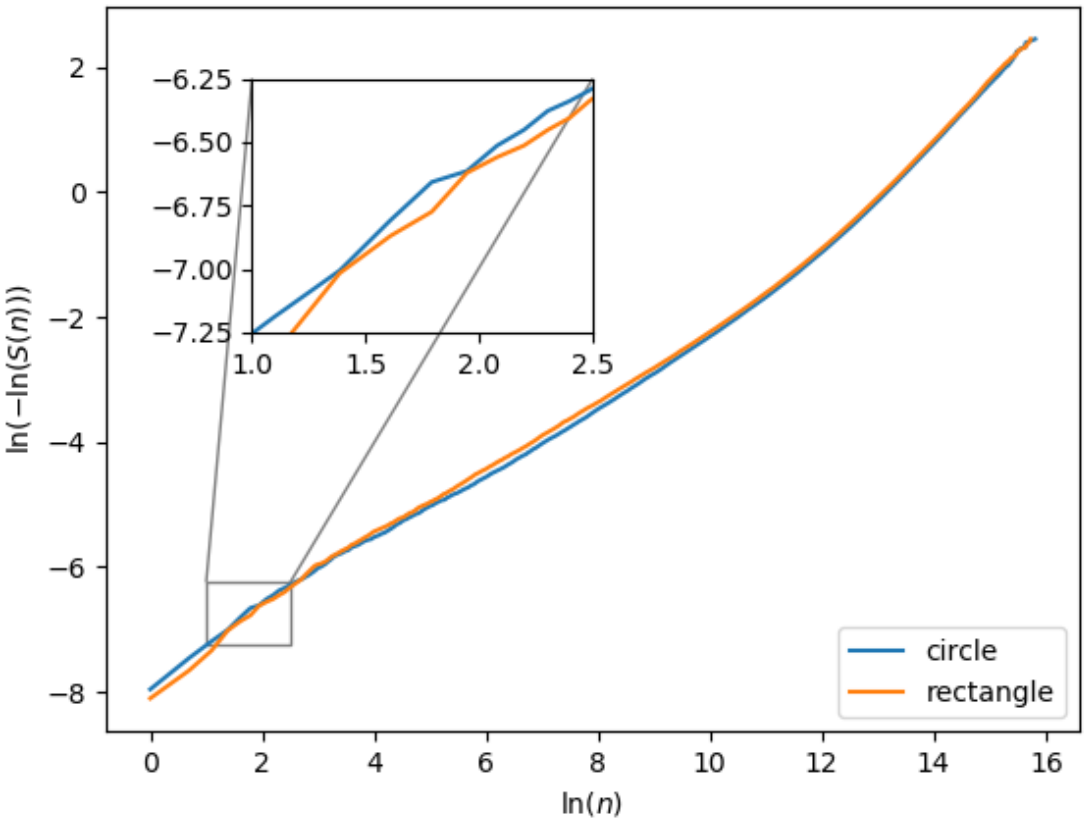
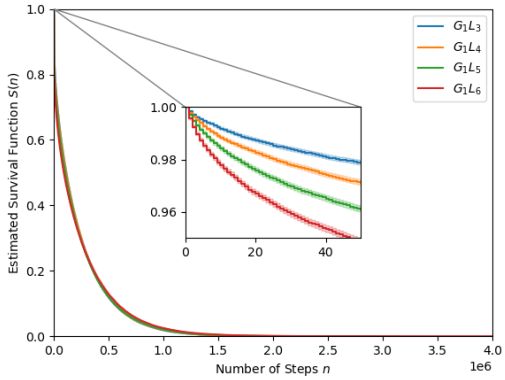
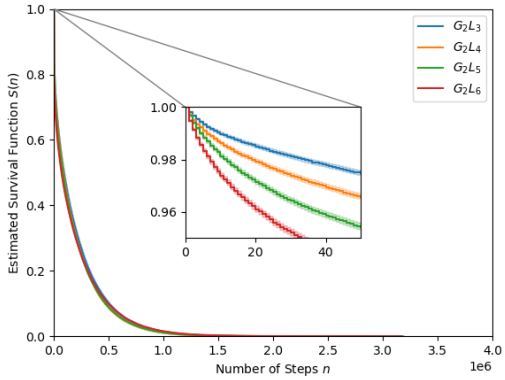


Figure 3.2: This is a graphical method for checking proportionality by looking for parallelism. As shown in the inset plot, two curves cross at some points and their shapes vary over time. Moreover, $p < 0.05$ in the non-proportional test. Thus, the survival data for circle and rectangle does not satisfy the proportional hazard assumption.



(a)



(b)

Figure 3.3: (a) and (b) are survival functions for branching structures in G_1 and G_2, respectively. n is the number of steps taken by the particle from the initial to the stop pixel in LRWs.

3.2 Complicated Branching Structures

In the preceding section, the research hypothesis has been tested by the comparison of simple shapes. In the preceding section, the research hypothesis has been tested by the comparison of simple shapes. More complex branching structures, as shown in Fig. C.2 and Fig. C.3, are produced in this section for further test and understanding of the survival curve. The branching structures in the images are equal-area and vertically symmetric, but the template in G_1 is shorter and narrower than in G_2. In G_i, $i = 1, 2$, the larger number of iterations j, the more complicated structures L_j, $j = 3, 4, 5, 6$, with more nodes.

3.2.1 Output Analysis of $S(n)$

The inset plot in Fig. 3.3 shows that the decay rates of $S(n)$ for L_j, $j = 3, ..., 6$, are significantly distinct. The graphical representation of short time behaviours of survival functions is consistent with the analytical result since bigger j results in the larger perimeter of the branching structure.

As shown in Fig. 3.4, the survival function is divided into several coloured segments. In Fig. 3.5, Fig. 3.6, Fig. 3.7, Fig. 3.7, and Fig. 3.9, particles' initial and stop positions are visualized by the scatter plots with the same colour as the segment to understand the underlying stochastic process and its properties. Moreover, the black region in the initial position plot is the target branching structure.

In reality, LRWs is latent and cannot be observed directly. However, those scatter plots are the visual representations of revealing the first-passage properties of particles in LRWs. For example, particles will be absorbed in a shorter time if their initial positions are closer to the target because they have less randomness. Moreover, the broader and longer the in-between space of branches is, the more particles are generated with a larger variance for the number of steps. Therefore, each segment of the survival curve carries massive geometric and spatial information about the unoccupied area of the binary image (i.e. the region with black

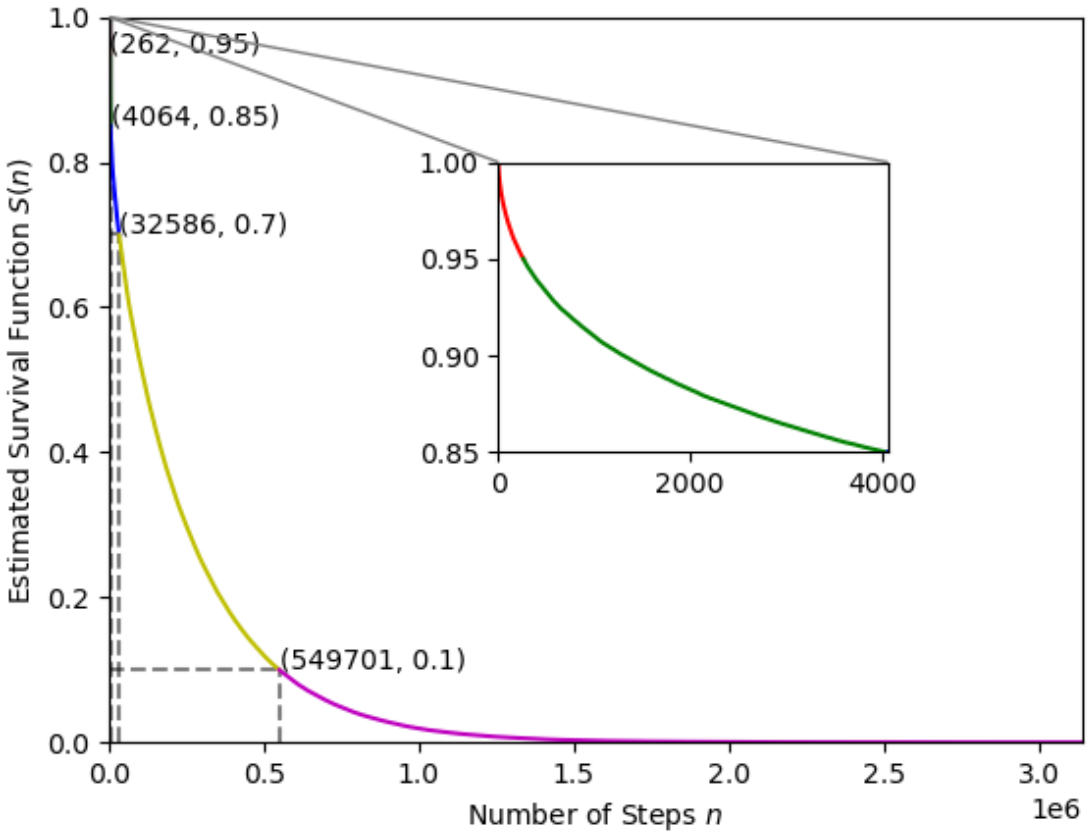


Figure 3.4: It is the estimated survival function for LRWs in $G_1 L_3$.

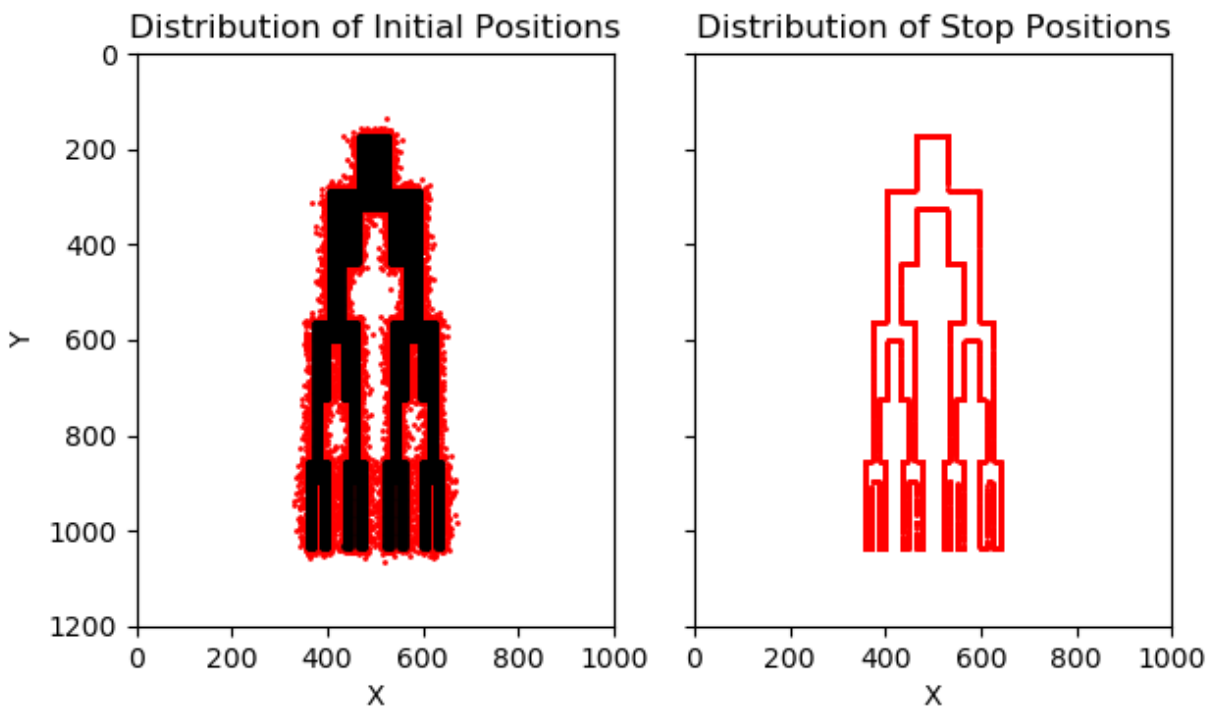


Figure 3.5: 5% of particles in the LRWs coloured by red will be absorbed within 262 steps. The left subfigure tells us that their initial positions are distributed in the space closely surrounding the target branching structure. The right subfigure shows that the red particles describe the entire boundary of the object.

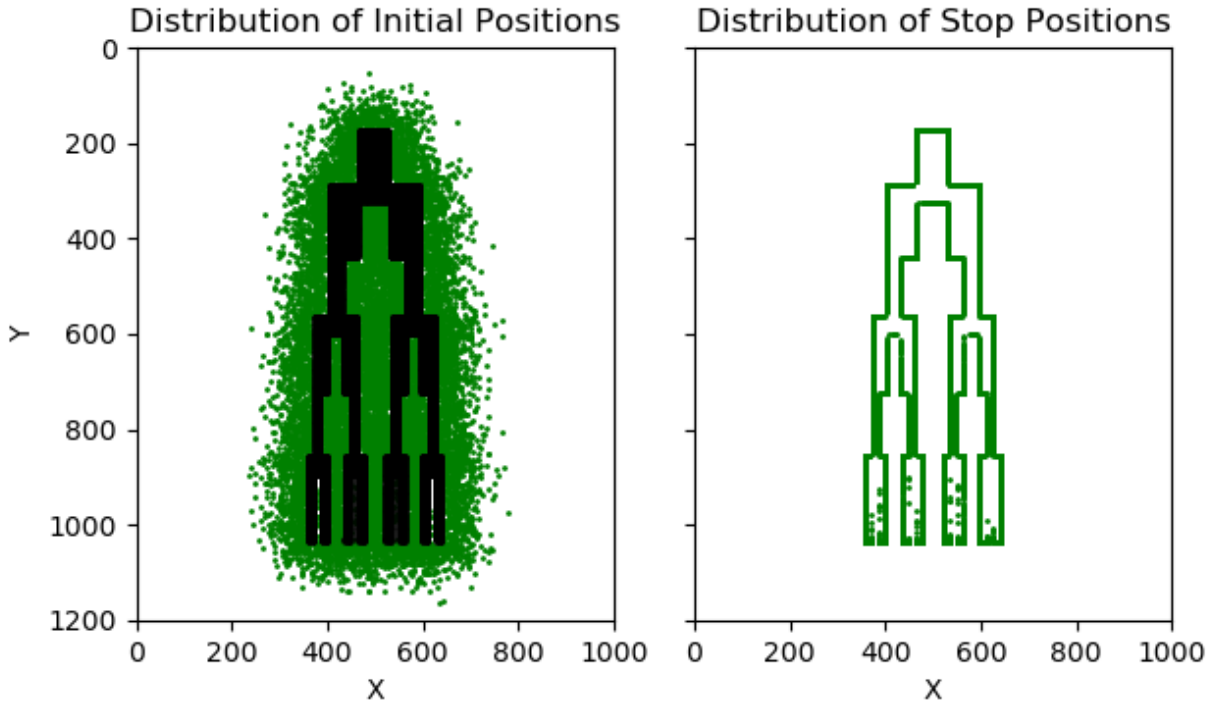


Figure 3.6: Around 10 percent of particles are near to the object, whose steps ranged from 262 to 4064. Compared with Fig. 3.5, the green points also dispersed thoroughly between the branches. Nevertheless, some in-between area of the bottom limbs is empty. From the right plot, we can tell that particles can not characterize the whole boundary of the object since they can not encounter some parts of the edge of the bottom branches.

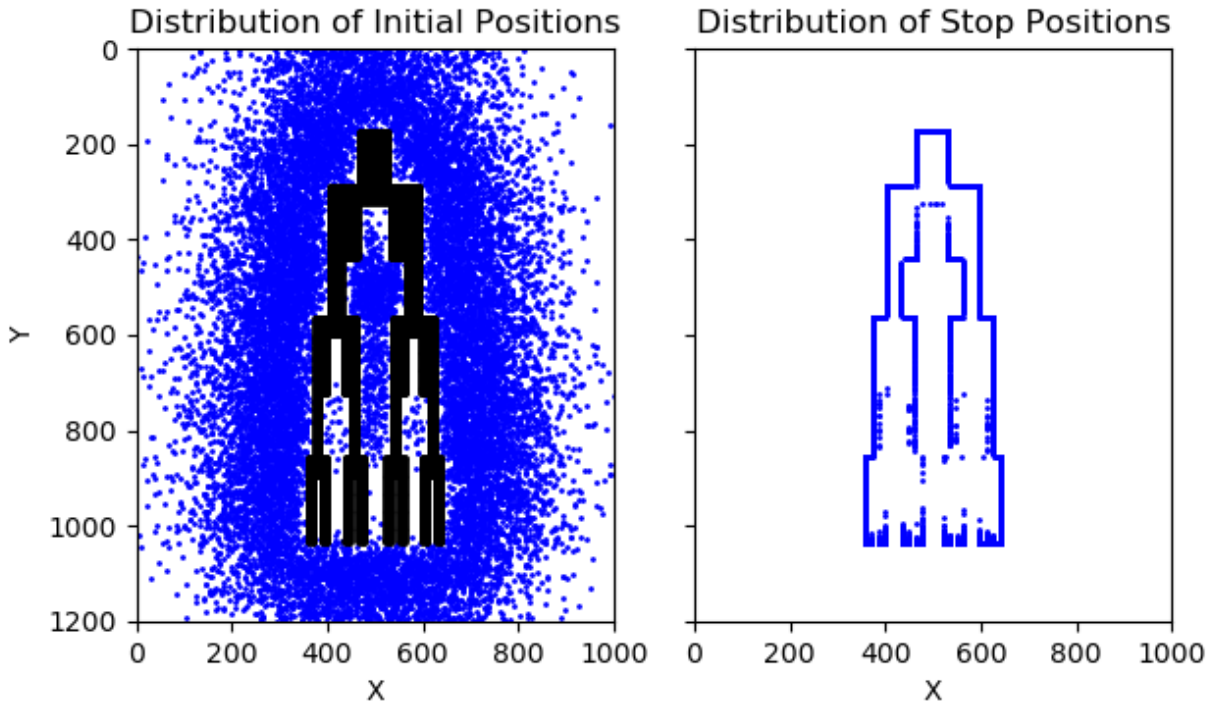


Figure 3.7: Approximate 15 percent of particles originally started LRWs from the region a little bit far away from the target object. The number of steps for the blue particles is from 4064 to 32586. Not like the Fig. 3.6, fewer particles are located in the area in-between the narrower branches. Moreover, if the initial sites of particles are further from the branching structure's external boundary, they will be more dispersive. Only some upper parts of the internal border can be depicted by the blue particles in the right subfigure.

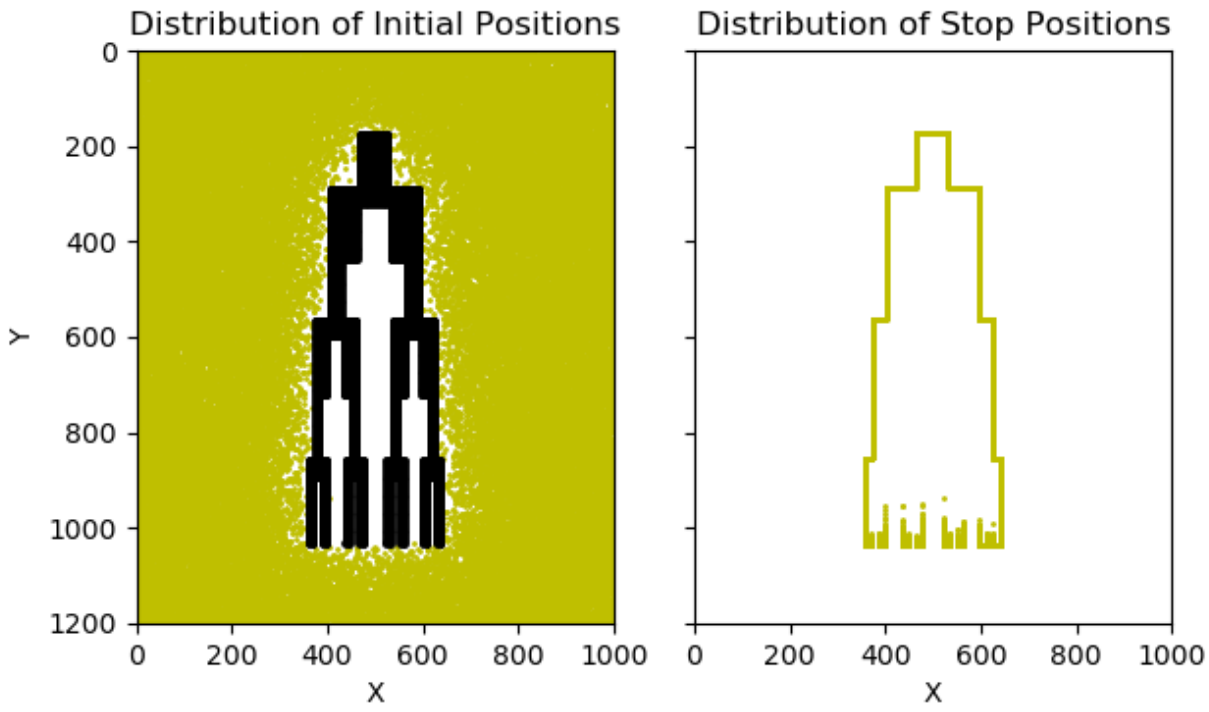


Figure 3.8: The left subplot displays the initial positions of 60% of particles in LRWs, which are distributed uniformly around the external outline of the branching structure. Moreover, none of them initially started walking from the space in-between the branches. Compared with the Fig. 3.7, the right subplot shows the object's entire external boundary and some internal border of its terminal limbs.

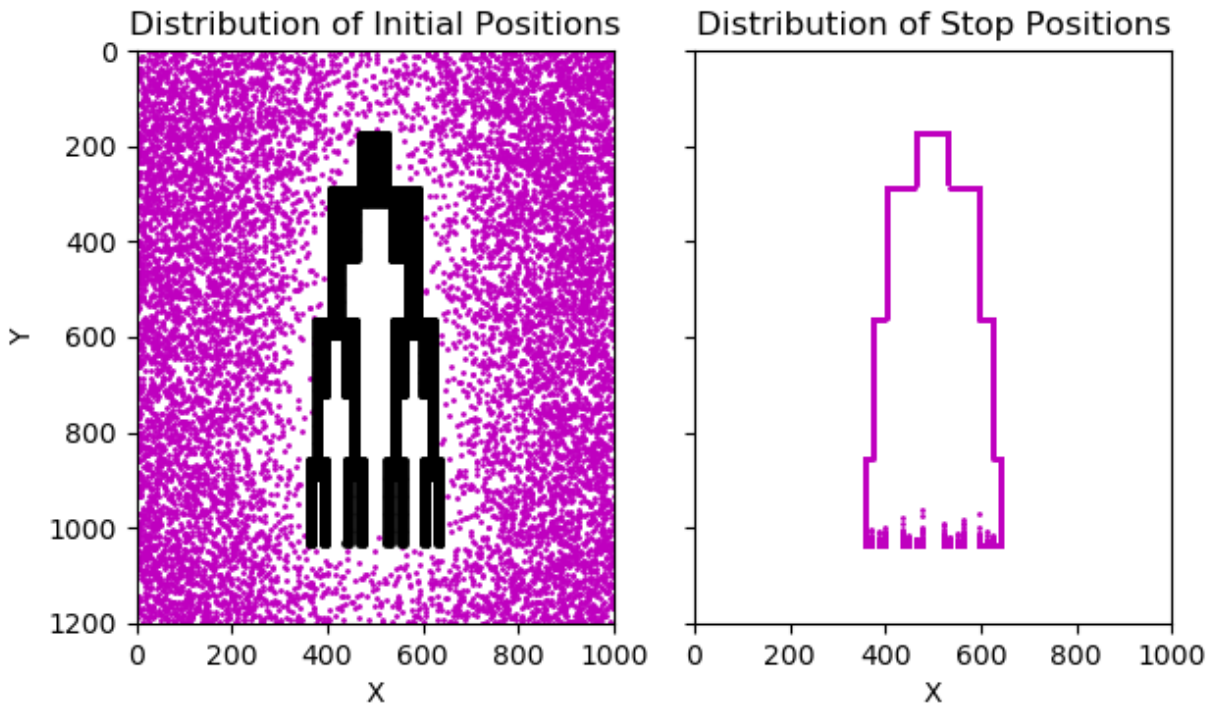


Figure 3.9: Only 10 percent of particles will still survive when their number of steps is bigger than 549701. They are distributed initially from a region far away from the object and close to the image's edges. Similar to Fig. 3.8, pink particles cannot delineate too much internal boundary of the branching structure.

		p			
		Log-rank	TW	GB	FH
G_1 L_3	G_1 L_4	0.4393	0.0285	0.0005	0.0005
	G_1 L_5	0.0	0.0	0.0	0.0
	G_1 L_6	0.0	0.0	0.0	0.0
G_1 L_4	G_1 L_5	0.0007	0.0	0.0	0.0
	G_1 L_6	0.0002	0.0	0.0	0.0
G_1 L_5	G_1 L_6	0.7223	0.0	0.0	0.0

Table 3.2: The differences between the pairwise survival functions for branching objects in G_1 are statistically significant.

		p			
		Log-rank	TW	GB	FH
G_2 L_3	G_2 L_4	0.0	0.0	0.0	0.0
	G_2 L_5	0.0	0.0	0.0	0.0
	G_2 L_6	0.0	0.0	0.0	0.0
G_2 L_4	G_2 L_5	0.0016	0.0	0.0	0.0
	G_2 L_6	0.0004	0.0	0.0	0.0
G_2 L_5	G_2 L_6	0.7199	0.0	0.0	0.0

Table 3.3: As mentioned before, the log-rank test will lose power if the proportional hazard assumption is violated. Except for the log-rank test, other statistical tests' results tell us that the pairwise survival functions are statistically different.

pixels) and the boundary of a simply connected domain (i.e. the artificial branching structure with white pixels).

Some weighted log-rank tests can be utilized to detect the early or late differences between the pairwise overlapping or crossing survival curves. In Table 3.2 and Table 3.3, TW is the abbreviation for Tarone-Ware test, GB is for Gehan-Breslow test, and FH is for Fleming-Harrington test. However, p values in the tables are not too informative, and the log-rank test under conditions of non-proportional hazards leads to misleading results. Hence, distance measures are alternative methodologies for quantifying the discrepancy between survival functions.

It is assumed that $\widehat{S}_1(t)$ and $\widehat{S}_2(t)$ are the Kaplan-Meier estimators of the survival functions for random variable $T_1 > 0$ and $T_2 > 0$, respectively. Let $(\tau_j)_{j=1,2,\ldots,N}$ are distinct increasing observed times when the event of interest take place.

In the vector space, the distance between $\widehat{S}_1(t)$ and $\widehat{S}_2(t)$ can be defined as the L_p norm of their difference,

$1 \leq p \leq \infty$, where

$$
\begin{aligned}
d_p(\widehat{S}_1(t), \widehat{S}_2(t)) &= \|\widehat{S}_1(t) - \widehat{S}_2(t)\|_p \\
&= \Big(\sum_{j=1}^{N} |\widehat{S}_1(\tau_j) - \widehat{S}_2(\tau_j)|^p \Big)^{\frac{1}{p}} \\
&= \Big(\sum_{j=1}^{N} |P(T_1 > \tau_j) - P(T_2 > \tau_j)|^p \Big)^{\frac{1}{p}}
\end{aligned}
\tag{3.1}
$$

As shown in Fig. 3.10, $p = 1$ in Eq. 3.1, the dissimilarities between each pair of survival functions are measured by d_1 and depicted by colors in the heat map. The cells on the main diagonal are white because the distance of an object from itself is zero. The off-diagonal cells are symmetric, which are darker indicating more significant dissimilarities between survival functions or curves. In other words, the color of cells represents the variation in the shapes of branching structures in the artificial images.

The top left and bottom right 4×4 cells are the in-group shape comparison of the branching structures. For each column, the cells become darker gradually since the objects in each group are more complicated as j increase. The pale yellow cells adjacent to the diagonal, including (G_1L3, G_1L4), (G_1L4, G_1L5), $(G_1L5, G_1L6))$, (G_1L3, G_1L4), (G_1L4, G_1L5), and $(G_1L5, G_1L6))$, implies that the morphological changes are not dramatically different.

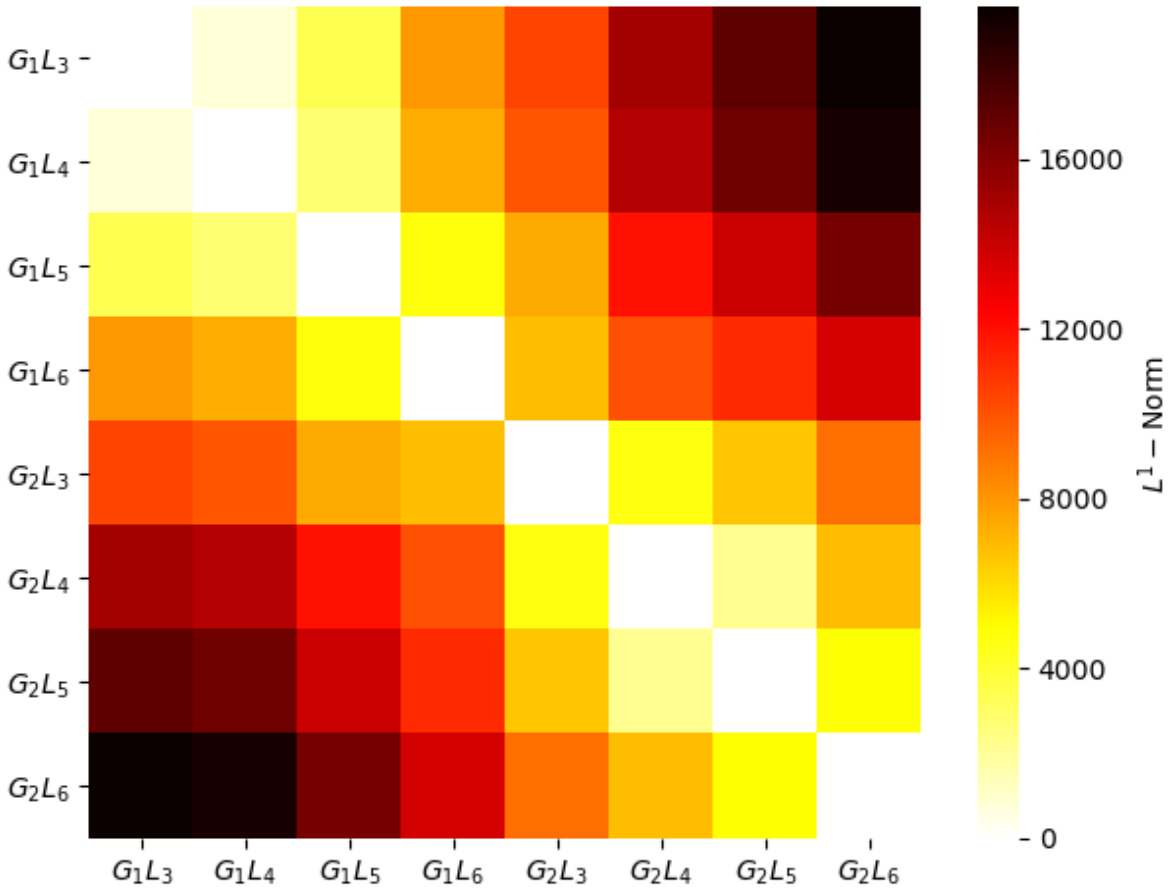


Figure 3.10: The distance matrix of pairwise survival functions for artificial images is visualized by a heat map, where L^1 norm is the distance metric.

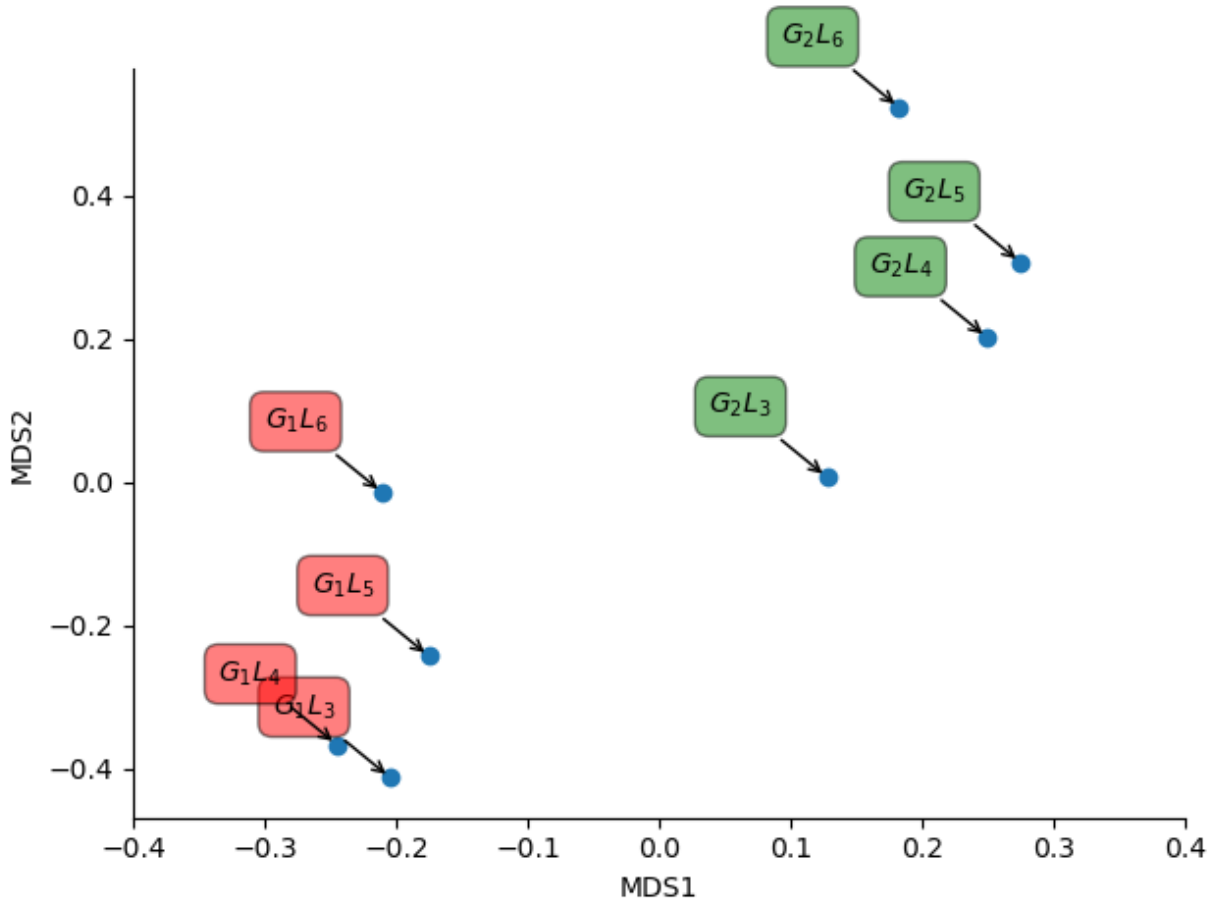
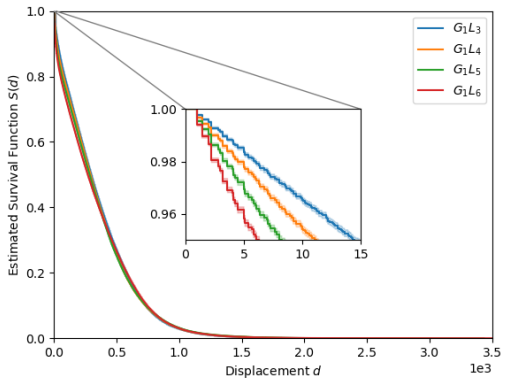
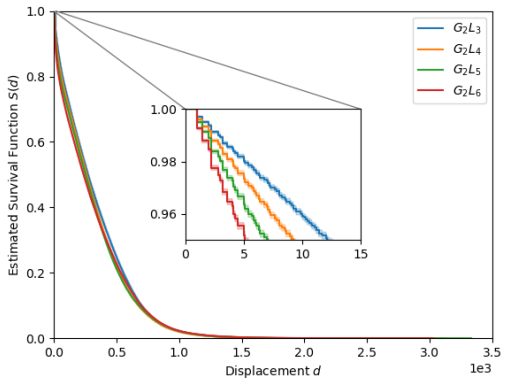


Figure 3.11

3.2.1.1 $S(d)$



(a)



(b)

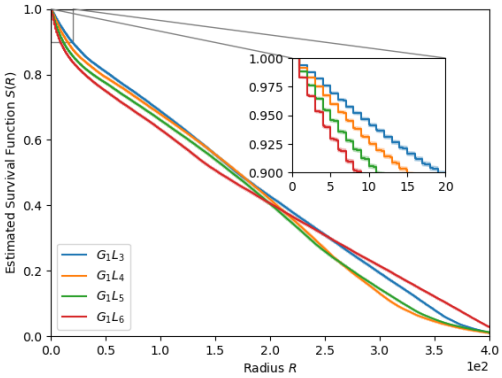
Figure 3.12

		p			
		Logrank	TW	GB	FH
$G_1\ L_3$	$G_1\ L_4$	0.0	0.0	0.0	0.0
	$G_1\ L_5$	0.0	0.0	0.0	0.0
	$G_1\ L_6$	0.0	0.0	0.0	0.0
$G_1\ L_4$	$G_1\ L_5$	0.0072	0.0	0.0	0.0
	$G_1\ L_6$	0.0003	0.0	0.0	0.0
$G_1\ L_5$	$G_1\ L_6$	0.2883	0.0	0.0	0.0

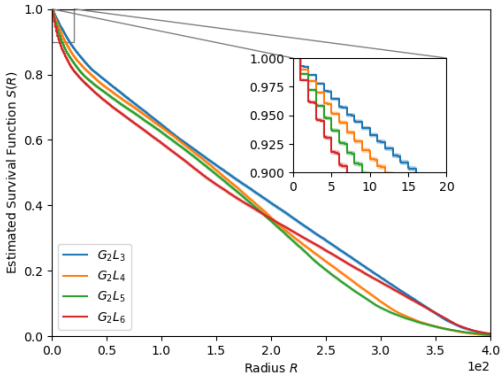
Table 3.4

		p			
		Logrank	TW	GB	FH
$G_2\ L_3$	$G_2\ L_4$	0.0	0.0	0.0	0.0
	$G_2\ L_5$	0.0	0.0	0.0	0.0
	$G_2\ L_6$	0.0	0.0	0.0	0.0
$G_2\ L_4$	$G_2\ L_5$	0.0001	0.0	0.0	0.0
	$G_2\ L_6$	0.0015	0.0	0.0	0.0
$G_2\ L_5$	$G_2\ L_6$	0.7019	0.0	0.0	0.0

Table 3.5



(a)



(b)

Figure 3.13

		p			
		Logrank	TW	GB	FH
$G_1\ L_3$	$G_1\ L_4$	0.0	0.0	0.0	0.0
	$G_1\ L_5$	0.0	0.0	0.0	0.0
	$G_1\ L_6$	0.0	0.0	0.0	0.0
$G_1\ L_4$	$G_1\ L_5$	0.1773	0.0	0.0	0.0
	$G_1\ L_6$	0.0	0.0	0.0	0.0
$G_1\ L_5$	$G_1\ L_6$	0.0	0.0	0.0	0.0

Table 3.6

3.2.1.2 $S(R)$

		p			
		Logrank	TW	GB	FH
$G_2\ L_3$	$G_2\ L_4$	0.0	0.0	0.0	0.0
	$G_2\ L_5$	0.0	0.0	0.0	0.0
	$G_2\ L_6$	0.0	0.0	0.0	0.0
$G_2\ L_4$	$G_2\ L_5$	0.0	0.0	0.0	0.0
	$G_2\ L_6$	0.0	0.0	0.0	0.0
$G_2\ L_5$	$G_2\ L_6$	0.0	0.0	0.0253	0.0253

Table 3.7

3.2.2 Conclusion

- In a short time, the survival function of rectangle decays faster than the circle, which conforms to the analytical results.
- The differences of estimated survival functions between circle and rectangle are statistically significant, which coincides with the real shape dissimilarities.
- Within a same group, when t is small, the more branching the object is, the faster the survival function decays.
- Within a same group, the pairwise survival functions are statistically different.
- The corresponding target structures in G_1 and G_3 are invariant shapes under translation since their survival function are not statistically different. In other words, periodic boundary conditions of the image can eliminate the effect of the locations.
- LRWs can describe and classify the geometries, their spatial configurations, and the unoccupied area in the image.

LRWs in Real Root Images

Conclusion

Future Work

Appendix A

Numerical Methods for Solving Parabolic Partial Differential Equations

A.1 Introduction

A.2 Summary of Commonly Used Numerical Techniques

A.3 Limitation in Practice

Appendix B

Method Validation in Annulus

B.1 Analytical Results

B.1.1 Shape Description

- Problem domain Ω: the region bounded by two concentric circles
- Radius of the larger circle: b
- Radius of the smaller circle: a

B.1.2 Solving Initial-Boundary Value Problem (IBVP)

B.1.2.1 Methods

- Dimensional Analysis: non-dimensional variables
  - $\mu = \frac{b}{a}$
  - $\tau = \frac{t}{a^2}$
  - $\hat{r} = \frac{r}{a}$
- Method of separation of variables

B.1.2.2 Mathematical Equations

- Diffusion equation

$$u_\tau = (u_{\hat{r}\hat{r}} + \frac{1}{\hat{r}} u_{\hat{r}} + \frac{1}{\hat{r}^2} u_{\theta\theta}) \tag{B.1}$$

- Uniform initial condition

$$u(\hat{r}, \theta, 0) = \frac{1}{|\Omega|} \tag{B.2}$$

- Homogenous Dirichlet B.C.

$$u(1, \theta, \tau) = 0 \tag{B.3}$$

- Homogenous Neumann B.C.

$$\hat{r} u'(\mu, \theta, \tau) = 0 \tag{B.4}$$

B.1.2.3 Heat Content Calculation

$$S(\tau) = \int_0^{2\pi} d\theta \int_1^{\mu} \hat{r} d\hat{r} u(\hat{r}, \theta, \tau) \tag{B.5}$$

B.2 Numerical Approximation

B.2.1 Eigenvalues $\lambda_{0,n}$

- Properties
  - $\lambda_{0,n} \in \mathbb{R}^+$, $(n \in \mathbb{N}_+)$
  - Monotonicity and Periodicity
- Estimation
  - $\lambda_{0,n} \in ((n-1)\pi, (n+1)\pi)$ [1]
  - Bisection method [8]

B.2.2 Approximation of $u(\hat{r}, \theta, \tau)$ and $S(\tau)$

- Direct summation
- Series acceleration methods

B.3 Comparison of Numerical and Analytical Results

B.3.1 Sample Size Evaluation

B.3.2 Comparison of $S(\tau)$ and $S(n)$

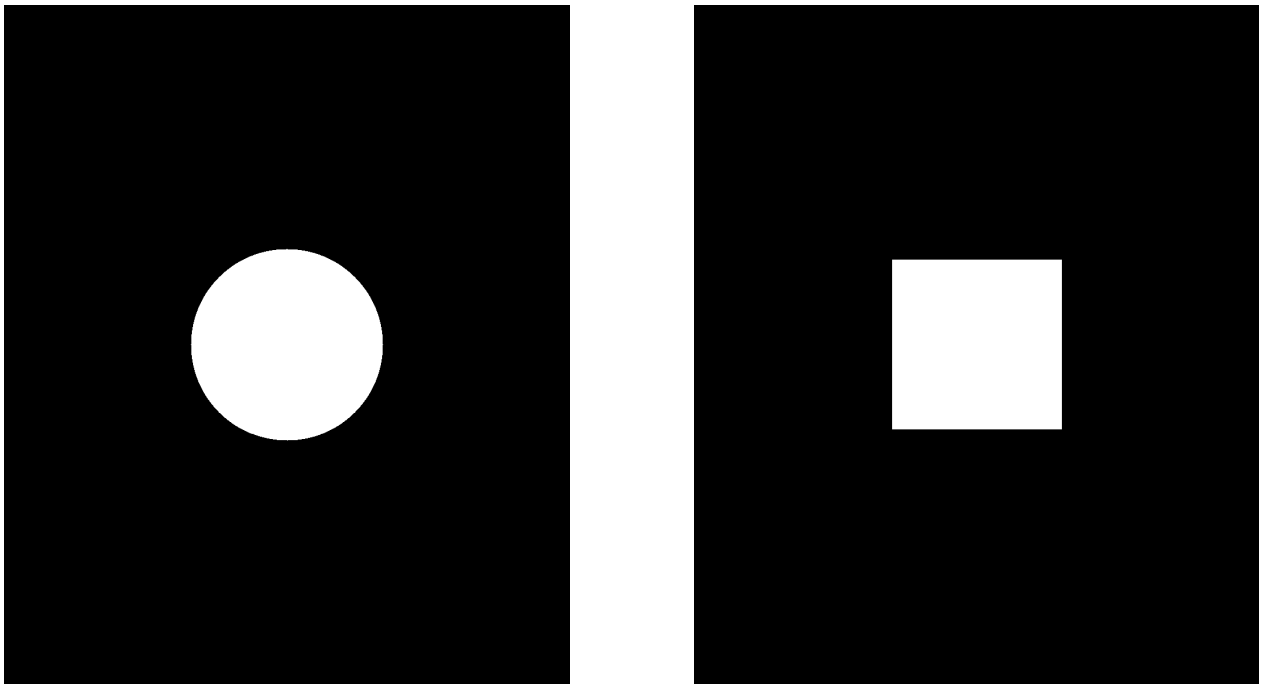
B.4 Conclusion

- The estimated survival function of LRWs is consistent with the analytical result.
- The number of particles in LRWs determined by DKW inequality is large enough to generate reproducible statistical results.

Appendix C

Artificial Images

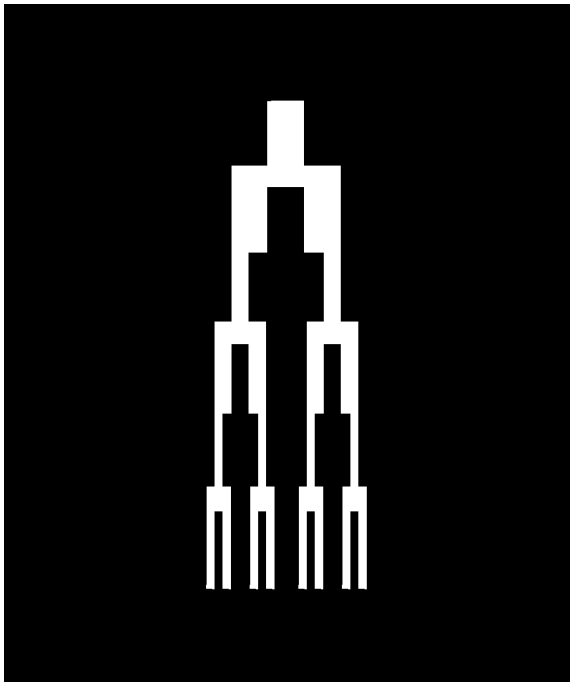
C.1 Simple Shapes



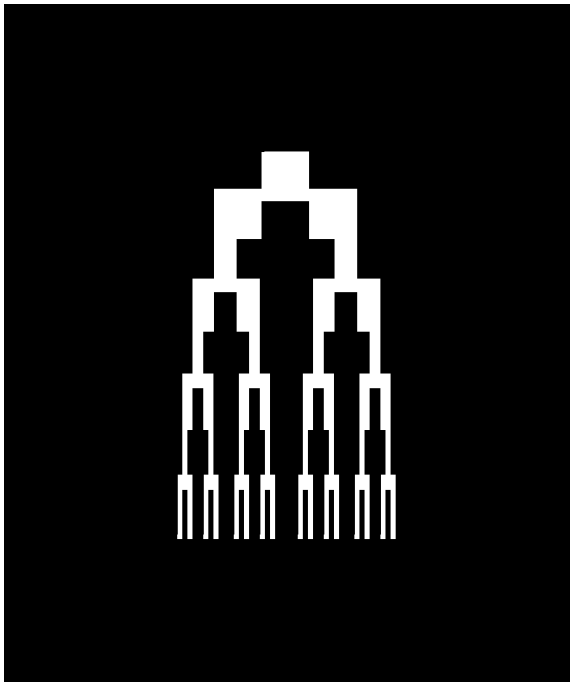
(a) Circle **(b)** Rectangle

Figure C.1: Each image size is 1200 by 1000 pixels with 90000 white pixels.

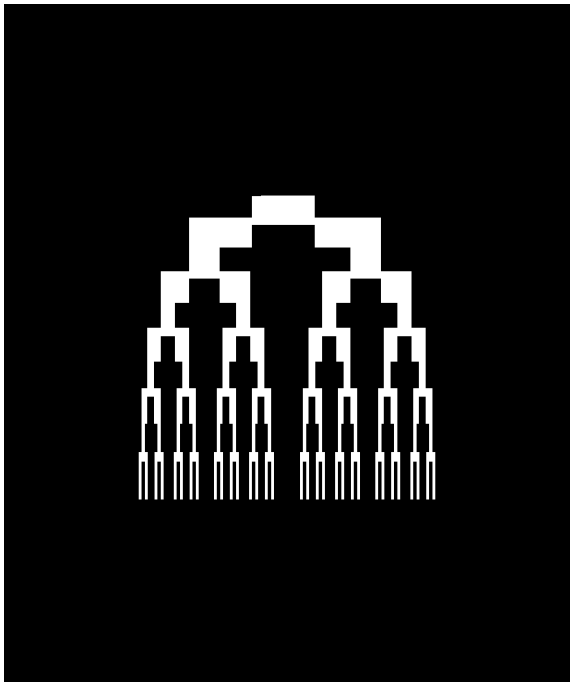
C.2 Complicated Branching Structures



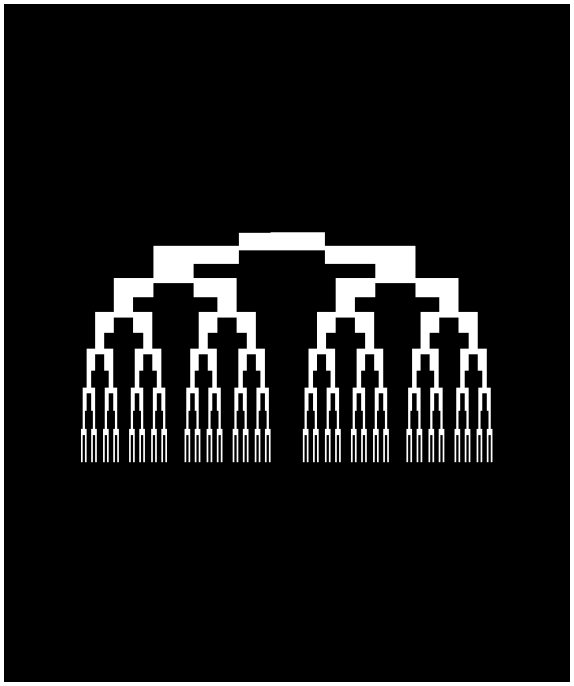
(a) G_1L_3



(b) G_1L_4

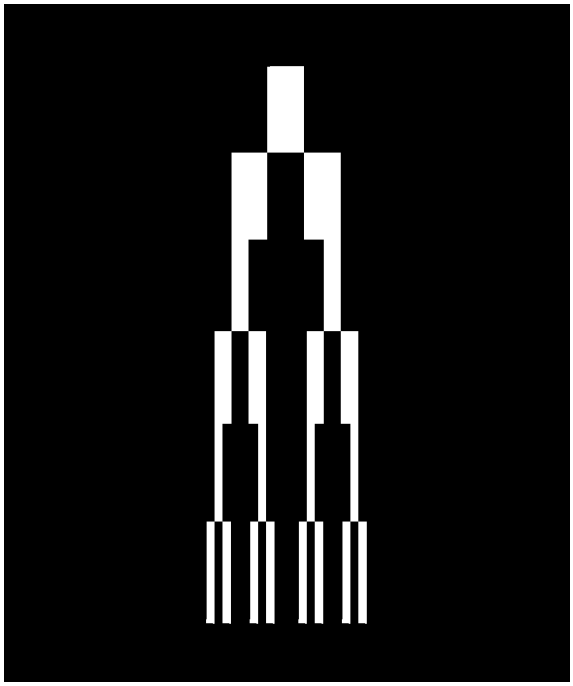


(c) G_1L_5

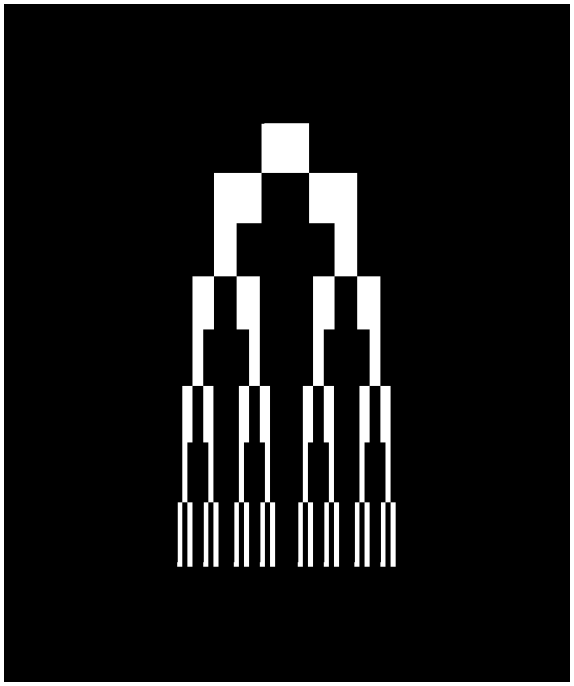


(d) G_1L_6

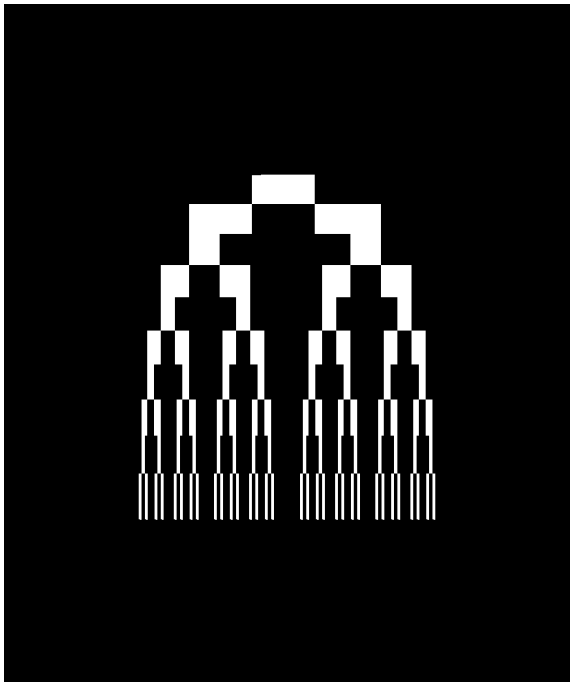
Figure C.2: In the group one, each image size is 1200 by 1000 pixels with 90000 white pixels.



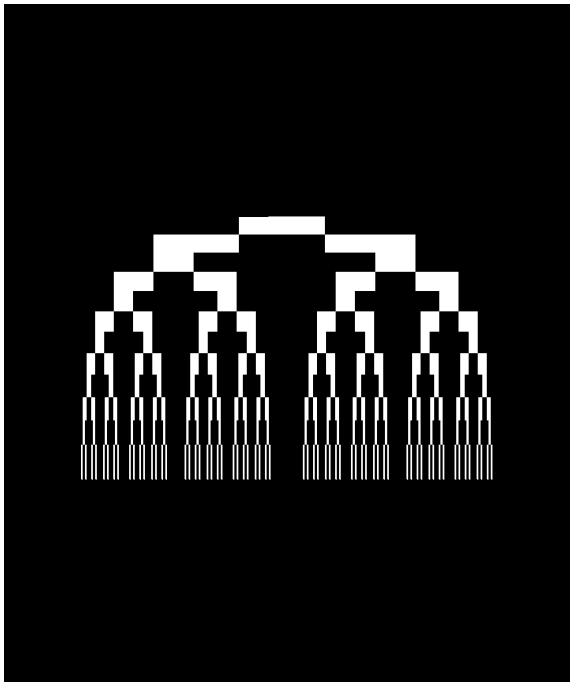
(a) G_2L_3



(b) G_2L_4



(c) G_2L_5



(d) G_2L_6

Figure C.3: In the group two, each image size is 1200 by 1000 pixels with 90000 white pixels.

References

[1] *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.26 of 2020-03-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[2] HPW Gottlieb. Hearing the shape of an annular drum. *The ANZIAM Journal*, 24(4):435–438, 1983.

[3] HPW Gottlieb. Eigenvalues of the laplacian with neumann boundary conditions. *The ANZIAM Journal*, 26(3):293–309, 1985.

[4] Daniel Grieser and Svenja Maronna. Hearing the shape of a triangle. *Notices of the American Mathematical Society*, 60(11):1440–1447, 2013.

[5] Mark Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4P2):1–23, 1966.

[6] Mohamed A Khabou, Lotfi Hermi, and Mohamed Ben Hadj Rhouma. Shape recognition using eigenvalues of the dirichlet laplacian. *Pattern Recognition*, 40(1):141–153, 2007.

[7] BD Sleeman and EME Zayed. Trace formulae for the eigenvalues of the laplacian. *Zeitschrift für angewandte Mathematik und Physik*, 35(1):106–115, 1984.

[8] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake Vand erPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[9] EME Zayed. Heat equation for an arbitrary doubly-connected region in R^2 with mixed boundary conditions. *Zeitschrift für angewandte Mathematik und Physik*, 40(3):339–355, 1989.