Gathering of Agents on a Line

Dmitry Rabinovich
Gathering of Agents on a Line

Research Thesis

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science

Dmitry Rabinovich

Submitted to the Senate of the Technion - Israel Institute of Technology
Kislev 5778, Haifa, December 2017
I want to dedicate this work to my loving wife Irina and my children Maayan, Rinjin and April Christine.
I would first like to thank my thesis advisor Prof. Alfred M. Bruckstein in the Faculty of Computer Science at the Technion. The door to Prof. Bruckstein office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this thesis to be my own work, but steered me in the right direction whenever he thought I needed it.

I would also like to thank my fellow Antz who were involved in numerous discussions, group meetings, lengthy office brain storms for this research project: Michael Amir, Ariel Barel, David Dovrat, Roece Francos, Rotem Manor and Ilana Segal. Without their passionate participation and input, the research could not have been successfully conducted.

Finally, I must express my very profound gratitude to my parents, my brother and to my girlfriend for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Thank you,

Dmitry Rabinovich

The Generous Financial Help of the Technion is Gratefully Acknowledged
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>List of Symbols and Notations</td>
<td>xi</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2 Problem definition</td>
<td>7</td>
</tr>
<tr>
<td>3 Mathematical modeling</td>
<td>9</td>
</tr>
<tr>
<td>4 Convergence of the Inner points</td>
<td>13</td>
</tr>
<tr>
<td>4.1 A return to the original model</td>
<td>17</td>
</tr>
<tr>
<td>5 Bounding the outer point distributions</td>
<td>19</td>
</tr>
<tr>
<td>6 Convergence to stationary distribution of the outer points</td>
<td>23</td>
</tr>
<tr>
<td>6.1 Markov Chain over ( X_d ) and its properties</td>
<td>24</td>
</tr>
<tr>
<td>6.2 Transition probability matrix’s properties for reversible Markov chain</td>
<td>25</td>
</tr>
<tr>
<td>6.3 Cheeger’s inequality</td>
<td>26</td>
</tr>
<tr>
<td>6.4 Applying Cheeger’s inequality</td>
<td>27</td>
</tr>
<tr>
<td>7 Simulations and results</td>
<td>37</td>
</tr>
<tr>
<td>8 Conclusions</td>
<td>41</td>
</tr>
<tr>
<td>Appendices</td>
<td>43</td>
</tr>
<tr>
<td>A Markov Chains: Basic Theory</td>
<td>45</td>
</tr>
<tr>
<td>A.1 Definition and Examples</td>
<td>45</td>
</tr>
<tr>
<td>A.2 Chapman-Kolmogorov Equations and the Transition Probability Matrix</td>
<td>45</td>
</tr>
<tr>
<td>A.3 Accessibility and Communicating Classes</td>
<td>46</td>
</tr>
<tr>
<td>A.4 Finite State Markov Chains</td>
<td>47</td>
</tr>
<tr>
<td>A.5 Stopping Times, Strong Markov Property</td>
<td>50</td>
</tr>
<tr>
<td>A.6 Countable state space case</td>
<td>51</td>
</tr>
<tr>
<td>B From real coordinates to integer model</td>
<td>53</td>
</tr>
<tr>
<td>C Random Walks and Hitting Times</td>
<td>57</td>
</tr>
<tr>
<td>Bibliography</td>
<td>61</td>
</tr>
<tr>
<td>Bibliography</td>
<td>61</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sensor and heading geometry for agent $v_i$</td>
</tr>
<tr>
<td>2</td>
<td>$v_1$ picked direction with agents on both sides</td>
</tr>
<tr>
<td>3</td>
<td>$v_1$ picked direction with agents on one side only and heading toward other agents</td>
</tr>
<tr>
<td>4</td>
<td>$v_1$ picked direction with agents on one side only and heading in the other direction</td>
</tr>
<tr>
<td>5</td>
<td>All agents with picked directions, and system’s Convex Hull. Thick headings are in allowed direction</td>
</tr>
<tr>
<td>6</td>
<td>6 agent system run. Dotted lines are historic Convex Hulls.</td>
</tr>
<tr>
<td>7</td>
<td>6 agent system run. Maximum pairwise distance plotted vs iterations.</td>
</tr>
<tr>
<td>8</td>
<td>12 agent system run. Dotted lines are historic Convex Hulls.</td>
</tr>
<tr>
<td>9</td>
<td>12 agent system run. Maximum pairwise distance plotted vs iterations.</td>
</tr>
<tr>
<td>10</td>
<td>$v_1$ and $v_2$ consistently pick &quot;bad&quot; directions</td>
</tr>
<tr>
<td>11</td>
<td>Close up on $v_2$. &quot;Bad&quot; directions are emphasized.</td>
</tr>
<tr>
<td>12</td>
<td>Typical simulation of 1D model over the course of few ticks. Red points will jump at next time tick. Pink points depicts the positions of red points in the preceding time tick.</td>
</tr>
<tr>
<td>13</td>
<td>Possible constellation of few points.</td>
</tr>
<tr>
<td>14</td>
<td>RP &quot;pushing&quot; inside point from bin 1 to bin 0.</td>
</tr>
<tr>
<td>15</td>
<td>Point relabeling after jump. LP &quot;pushed&quot; point to the right and RP jumped to the left.</td>
</tr>
<tr>
<td>16</td>
<td>LP &quot;pushed&quot; 0.</td>
</tr>
<tr>
<td>17</td>
<td>RP &quot;pushed&quot; 0.</td>
</tr>
<tr>
<td>18</td>
<td>LP jumps inside.</td>
</tr>
<tr>
<td>19</td>
<td>LP over-jumps inner points.</td>
</tr>
<tr>
<td>20</td>
<td>Some possible inner point constellations. Number pair denotes a number of inner points in left and right bin accordingly.</td>
</tr>
<tr>
<td>21</td>
<td>Simplified Markov chain associated with $Y$.</td>
</tr>
<tr>
<td>22</td>
<td>Finite Markov chain over $X_d$.</td>
</tr>
<tr>
<td>23</td>
<td>Markov chain associated with $\mathbb{R}^2$.</td>
</tr>
<tr>
<td>24</td>
<td>Markov chain associated with $\mathbb{R}^2_d$.</td>
</tr>
<tr>
<td>25</td>
<td>$\mathcal{S}_{k,j}$.</td>
</tr>
<tr>
<td>26</td>
<td>$\mathcal{S}_{k+2,j+2}$.</td>
</tr>
<tr>
<td>27</td>
<td>$\mathcal{S}<em>{k,j}$ and $\mathcal{S}</em>{j+2,m}$.</td>
</tr>
<tr>
<td>28</td>
<td>$\mathcal{S}_{k+2,m}$.</td>
</tr>
<tr>
<td>29</td>
<td>&quot;2&quot;-arm subchain.</td>
</tr>
<tr>
<td>30</td>
<td>&quot;3&quot;-arm subchain.</td>
</tr>
<tr>
<td>31</td>
<td>Two-arms subchain.</td>
</tr>
<tr>
<td>32</td>
<td>Convergence Time vs. Length of Initial Interval.</td>
</tr>
<tr>
<td>33</td>
<td>Convergence Time vs. Number of Agents.</td>
</tr>
</tbody>
</table>
Convergence Time with Best/Worst and Average bounds depicted. Best - is just pushing at each step.

Convergence Time vs. $\varepsilon$. Only $\varepsilon$'s < 0.5 lead to convergence in finite expected time.

Distribution of outer point distances from the bulk of inner points. In yellow distribution of total(between both outer points) distance is depicted. $\varepsilon = 0.1$.

Distribution of outer point distances from the bulk of inner points. In yellow distribution of total(between both outer points) distance is depicted. $\varepsilon = 0.4$.

Distribution of inner point constellations. $\varepsilon = 0.1$.

Distribution of inner point constellations. $\varepsilon = 0.4$.

LP/RP moved outside.

LP/RP moved inside.
List of Symbols and Notations

\[ <f, g >_{L^2(1/\pi)} = \sum_{x \in \mathcal{X}} f(x)g(x)/\pi(x) \]

\[ \chi(x) = ([x], \{x\}), \text{splitting the real number into its integer and fractional parts} \]

\[ \delta \]

length of a single jump, WLOG assumed to be 1

\[ \eta_t \]

is an \( t^{th} \) increment of \( \{X(t)\} \)

\[ \mathbb{E}(\cdot) \]

expectation of

\[ \hat{\delta} \]

same parity indicator of \( k \) and \( d \)

\[ \hat{h}(S_k) \]

value of \( h(S_k) \) adjusted to actual value of \( \hat{\delta} \)

\[ \lambda, \lambda_1 \]

eigenvalue of matrix, \( i^{th} \) eigenvalue of matrix where applicable

\[ \lambda^* \]

second largest eigenvalue

\[ \|\cdot\| \]

regular Euclidean norm on \( \mathbb{R} \)

\[ P \]

the transition probability matrix

\[ P_d \]

the transition probability matrix of Markov chain over state space \( \mathcal{X}_d \)

\[ \mathbb{R} \]

real line

\[ \mathbb{R}^n \]

real coordinate space in \( n \) dimensions

\[ \mathbb{Z}^n \]

\( n \) dimensional integer lattice

\[ W(t) \]

stochastic process, tracking the coordinate of RP

\[ X(t) \]

stochastic process, tracking the coordinate of RP, in case LP is absent

\[ Y(t) \]

the RP Markov chain, which bounds \( X(t) \)

\[ Z \]

the Markov chain for LP

\[ L_{II}(t) \]

the distance between the left-most and the right-most inner points in a continuous model

\[ L_I(t) \]

the distance between the left-most and the right-most inner points + 1

\[ \mathcal{M} \]

set of all functions from \( \mathcal{X} \) to \( \mathbb{C} \)

\[ \mathcal{S}, \mathcal{T} \]

subchains

\[ S_k \]

subchain \( S_{k,d} \) over \( \mathcal{X}_d \)

\[ S_{k,j} \]

continuous subchain in one of the arms, starting at node \( k \) and ending at node \( j \)
\[ T_{k,j} = S_k \cup S_j \]

- \( \mathcal{X} \): state space of Markov chain \( \mathcal{Y} \)
- \( \mathcal{X}_d \): finite subspace of state space \( \mathcal{X} \)
- \( S_t \): an unordered vector of agents' coordinates at time \( t \in \mathbb{N} \)
- \( \mu_0, \mu_k \): probability distributions of Markov chain over \( \mathcal{X} \) initial and at time \( k \)
- \( \mu_{0,d}, \mu_{k,d} \): probability distributions of Markov chain over \( \mathcal{X}_d \) initial and at time \( k \)
- \( \| \cdot \|_{L^2(1/\pi)} \): norm induced by \( <\cdot,\cdot>_{L^2(1/\pi)} \)
- \( \| \mu - \pi \|_{TV} \): Total variation norm of two probability distributions \( \pi \)
- \( \pi(\mathcal{Y}), \pi_d(\mathcal{Y}) \): stationary probability distribution of subspace \( \mathcal{Y} \)
- \( \pi(d) \): \( d^{th} \) entry of stationary distribution vector \( \pi \), with \( d \) in some index set
- \( \pi(x), \pi_d(x) \): stationary probability distribution of state \( x \)
- \( \pi_d \): stationary probability distribution of \( \mathbb{P}_d \)
- \( P(\cdot) \): probability measure
- \( \sigma_t = \begin{cases} 1, & \eta_t = 1 \\ -1, & \text{o.w.} \end{cases} \)
- \( \Upsilon = \sum_{i=1}^{S_r} \Upsilon_i \)
- \( \Upsilon_i \): time between \((i-1)^{th}\) and \(i^{th}\) pushes
- \( \epsilon \): probability of picking a wrong jump direction
- \( \xi_t \): random variable, denoting a Random Walk step at time \( t \)
- \( \{\cdot\} \): fractional part of a real number
- \( c = \frac{\epsilon}{1-\epsilon} \)
- \( C_m \): \( m^{th} \) Catalan number
- \( d(S_t) \): the maximum absolute difference of any two entries in \( S_t \)
- \( Dist(\mathcal{Y}, \mathcal{Z}) \): minimum distance between any pair of states of subspaces \( \mathcal{Y} \) and \( \mathcal{Z} \)
- \( h = \min_{\pi(\mathcal{Y}) \leq 1/2} \frac{Q(\mathcal{Y} \times \mathcal{Y}^c)}{\pi(\mathcal{Y})} \)
- \( I_A \): indicator random variable of event \( A \)
- \( n \): number of agents/points
- \( P(x, \mathcal{Y}) \): transition probability from state \( x \) to state subspace \( \mathcal{Y} \) in 1 step
- \( P(x, y), p(x, y) \): transition probability from state \( x \) to state \( y \) in 1 step
- \( p_k(x, y) \): probability to move from state \( x \) to state \( y \) in \( k \) steps
- \( Q(A, B) = \sum_{x \in A, y \in B} Q(x, y) \)
\( Q(x,y) = \pi(x)P(x,y) \)

\( S_C \)
the sum of all the initial points’ coordinate differences from the \textbf{left-most} inner point

\( T \)
stopping time

\( T_{\text{critical}} \)
in integer model \( Y \) the first time all the inner points gathered at the same location

\( T_{\text{push}}^k \)
\( k \text{th} \) push to the left in integer model \( Y \)

\( X \)
original model

\( X_k(t) \)
the position of \( k \text{th} \) agent at time \( t \) in original model, index \( k \) depends on the context

\( Y \)
integer model

\( Y_k(t) \)
the position of \( k \text{th} \) point in integer model at time \( t \)

2-arm subchain subchain of \( \mathbb{P}^2 \) over \( \mathbb{N}_+ \), consisting only of even states

3-arm subchain subchain of \( \mathbb{P}^2 \) over \( \mathbb{N}_+ \), consisting only of odd states, excl. 1

a.s. almost surely, with probability 1, WP 1

LP/RP current left/right-most point in the model

WLOG without loss of generality

WP \( p \) with probability \( p \)
Abstract

We consider a group of mobile agents on a line, identical and indistinguishable, memoryless, having the capability to only sense the presence of neighboring agents to the left and to the right. The agents’ rule of motion is as follows: at each moment, agents with neighbors on both sides stay put, while agents with neighbors on one side only jump with high probability \((1 - \varepsilon)\) a unit distance towards the neighbors (otherwise, with low probability \(\varepsilon\), they jump one unit away). We prove that all agents, except two, gather almost surely inside a unit size interval for any \(\varepsilon < \frac{1}{2}\) and do that in finite expected time. Two agents, the current left-most and right-most ones perform random walks biased towards the cluster of other agents. The cluster of gathered agents slowly moves on the line. Interesting interactions occur when the left and/or right Random Walkers reach the clustered agents and these interactions are completely analyzed herein. We give an upper bound estimate on the rate of convergence to stationary distribution of left and right Random Walkers, and test empirically our theoretical results.

We briefly examine the generalization of the model to higher dimensions and analyze some pathological cases. We conclude that some additional assumptions should be made, to apply our results to a more complex models.
Chapter 1

Introduction

For millennia humanity was amazed and amused by the complexity of the perceived cooperative behavior in colonies of ants, bees and termites, in flocks of birds, in school of fish, in swarms of locusts and even in crowds of people. While recognized as a widespread phenomenon in the nature, science advances of last couple of centuries only slightly enlightened us on the exact mechanism achieving these swarming behaviors. Social animals supply us an abundant variety of seemingly simplistic behavioral patterns giving rise to amazingly sophisticated coordinated problem solving solutions, see e.g. Okubo 1986, Camazine et al. 2002, Sumpter 2006, Garnier et al. 2007 and many more.

In robotics we are often interested in solving problems like patrolling a region, sweeping and cleaning a complex-shaped area, searching for stationary or moving objects in unfamiliar environment. While all such tasks could be handled in theory by highly non-trivial algorithms specifically designed for a high-end multi-sensor robotic systems with abundant processing power, we are interested in a quite a different approach. Inspired by the nature, we wish to understand the laws of inter-agent and agent-to-environment interactions, which leads to emergence of the desired cooperative behavior. Şahin in Şahin 2004 gave a definition and a basic motivation for the term swarm robotics, where complex system-level behavior emerges from the local interaction of the large number of relatively simple units(agents). Mohan and Ponnambalam in Mohan and Ponnambalam 2009 extensive review showed that practical considerations, and among them financial costs, fragility and complexity, in a meanwhile, led the field to seriously consider deployments of multiple interacting agents as an alternative to omni-potent, though not always perfect systems.

The scientific community, through many years of research set the focus on several important and basic challenges that should be addressed. Among other problems, the fundamental problem of gathering agents together, also known as clustering, geometric consensus or “distributed agreement” problem, emerged as a very important topic. Works in this area assume various types of agent motion laws, based on "how agent $i$ influences agent $j$". We should note a close resemblance to distributed computing where each agent is associated with a vertex in communication graph, and edge - communication link between vertexes $i$ and $j$ is added in case agent $i$ is a neighbor (i.e. sense the presence) of agent $j$. Each agent $i$ holds initially a value(s) $v_i$, and the goal for every agent is to compute a function of values $v_1, v_2, \ldots, v_n$, for example their average. Lynch 1996. The computation should be done as efficiently as possible by exchanging information over the communication channel. In classical models computing is cheap, and any problem could be solved locally, even NP-hard ones. On the other hand communication is prohibitively expensive. We take those models to even the extreme by abandoning vertex memory in between communication rounds.

Imagine many identical mobile agents dispersed in a space, being it on the plane, as for automobile vehicles, or in 3D, as for quad or multi-rotors. Our interest is to get all these mobile agents together in order to perform some task cooperatively as a cohesive group. Envisioned by readily available examples in nature, we require the agents' laws of motion to be very
simplistic, the same for everyone, i.e. no central management of such system, but at the same time lead to system convergence to a gathering point or at least to a small bounded region. Agents have a limited variety of senses: in some settings they could sense direction to other agents, sometimes, they are limited to sensing only presence of other agents in some limited neighborhood. We assume no direct inter-agent communication and also that agents do not posses a common frame of reference (i.e. direction to North) and have no capability to collect and store information about the environment and/or other agents in the system.

[Barel et al. 2016] and [Manor and Bruckstein 2016] gave comprehensive overviews of the Gathering problem, classifying all the models into continuous/discrete time settings, full/limited visibility of sensing, and the type of information gathered on neighboring agents, i.e. sensing of relative position(direction and distance) or bearing only(direction to neighbors). But is it possible to simplify our assumptions further? Is it possible to drop direction (and where appropriate distance) sensing? Or to exchange it to a much more crude one?

We consider the idea of [Barel et al. 2016]
and are interested in testing the properties of even simpler model of multi-agent system on the plane. As such we consider identical, oblivious agents capable of sensing only the presence of other agents, and constrained to move only if some half-plane is empty of other agents, otherwise agents remain still. We assume each agent has an unlimited visibility range, but the sensors are otherwise primitive and test for the presence of other agents in the half plane behind their heading (see Figure 1). At discrete time ticks each agent chooses its heading uniformly at random and decide

---

**Figure 1:** Sensor and heading geometry for agent $v_i$.

**Figure 2:** $v_1$ picked direction with agents on both sides

**Figure 3:** $v_1$ picked direction with agents on one side only and heading toward other agents
In Figures 2–5 possible states of the system are presented. We should note that only agents on the Convex Hull of the system can move, as any agent inside it is surrounded, and any direction it picks splits the group of agents in two non-empty parts.

Computer simulations (see Figures 6–9) showed that for any random initial arrangement of agents on the plane, after some time, the system "converges", i.e. gathers to a bounded disc.

In mathematical terms we could expect to find some "natural" Lyapunov function, which will ensure the convergence, but as shown in Figures 10 and 11 the system could possibly diverge even indefinitely under adverse conditions (i.e. choices of headings!). Hence we may conclude that such function cannot exist. Instead a probabilistic methods should be applied to prove almost sure convergence. This seemingly trivial problem has yet to be mathematically analyzed in a simple way. Hence, we are keen to find solutions to even simpler versions of the problem. One promising path, we believe, would be to analyze the behavior of agent locations’ projections on arbitrary lines.

We consider here a simple 1D problem. Our one dimensional problem preserves the same movement decision rule of each agent, specifically, agent senses other agents direction [it is now right or left only]. Each agent surrounded from both sides stay put, while agents, which are on the edge of the interval occupied by the agents, will move. To resemble the original two-dimensional problem, where agents could also move away from other agents, one-dimensional
agents are allowed to move away, but the probability of such event is set to be low [as it is in the original problem]. We shall denote it by $\varepsilon \in [0, 1)$. Moves are of length $\delta = 1$ to either left or right (other step sizes can be achieved by properly changing the scale of the system). Figure 12 presents a short run of the evolution of a 1D system with 10 agents randomly placed on the line.
Like in the original problem, only agents on Convex Hull of the system can move. In the 1D case - those are two extremal agents on the right and the left side. It is immediately clear that the distance between second left-most agent and second right-most agent is monotonically non-increasing. As can be seen from Figure 12 left- and right-most agents could change "identities". If points are labeled, and label is preserved between the discrete times when the synchronized motion occurs - different points will eventually move. We show, that the size of the interval, which contains all non-moving agents at time t, decreases to an interval of length $1(=\delta)$ for $\varepsilon < 1/2$ in a finite expected time, almost surely. The agents, which "switch" identities indefinitely, perform stochastic walks, closely resembling a bounded biased random walks. And, in time after all [currently] non-moving agents gathered in an interval of length 1, the probability to find moving agent at distance $d$ from this interval decreases exponentially with $d$. We further show that convergence to this probability measure is exponentially fast and give an upper bound on convergence time as a function of model parameters.

The structure of the thesis is: in Chapter 2 we present a problem we address, in Chapter 3 we simplify the problem and present the mathematical model used to prove the convergence of agents. In Chapters 4, 5 and 6 we make precise our convergence claims. We then conclude and discuss some possible future work.
Chapter 2

Problem definition

Given \( n \) points (agents) at some initial positions \( S_0 = \{x_1, x_2, \ldots, x_n\} \) on the real line \( \mathbb{R} \), and "bearing-only" sensing of other points (agents), each point can move due to the following rules:

**Motion laws**

- Agent \( k \) does not move if other points exist to its left and right.
- Otherwise, with probability \((1 - \varepsilon)\) it makes a \( \delta \)-step (assume, by rescaling and w.l.o.g., that \( \delta = 1 \)) in the direction of the other points, and with probability \( \varepsilon \), \( \delta \)-step away from them.

We assume that

**Discrete-Time Synchronous Assumption**

The system is discrete-time and synchronous, i.e. the agents check their status and moves at the same discrete times.

The motion of agent \( k \) at time \( t \) depends only on the location of the other points just before time \( t \).

We also assume, that

**Separation Assumption**

No two points are co-located or will be co-located in the future, as a result of the above-mentioned movement rules.

(This can be ensured if the fractional parts of the \( x_i \) differ!)

Assume \( S_t \) - is an unordered vector of point coordinates at time \( t \). Let each point be associated with its index in natural \( \mathbb{R} \)-ordering in \( S_0 \), i.e. point which was 3rd from the left in \( S_0 \) will be regarded as third throughout this chapter. Let \( d : \mathbb{R}^n \to \mathbb{R} \), given by

\[
d(S_t) = \max_{i,j} \|S_t(i) - S_t(j)\|
\]

be the diameter of \( S_t \).

We are eventually interested in estimating

\[
F(s, t | \varepsilon, \delta, S_0) \triangleq P(d(S_t) < s\delta) \quad \text{if } \delta = 1 \Rightarrow P(d(S_t) < s) \tag{2.1}
\]

If \( P(d(S_t) < s) \approx 1 \) for \( t \to \infty \) and \( s \leq 2, 3 \), then the points are deemed concentrated in a small sized interval. And this is a result we strive to obtain.
Chapter 3

Mathematical modeling

We show that our original problem, despite being defined on the Real line (\(\mathbb{R}\)), can be mapped to a Markov chain over \(\mathbb{Z}^n\).

We start by few basic observations on the problem. We assumed two points will never be co-located, hence

**Observation 3.1.** Precisely two points move at each time step.

The observation is trivially backed up by the fact, that only boundary points are candidates to move, and the one-dimensional ball has two points on the boundary, namely the left and the right ends of the interval containing all the points.

**Observation 3.2.** If we label arbitrary agent \(k\) with initial position \(X_k(0) = x_0\), then fractional part of this point location’s coordinate is the same forever, i.e. \(\{x_0\}\), where \(\{x\}\) defined as \(\{x\} = x - \lfloor x \rfloor\).

This is indeed the case, since the agent could only change its coordinate by one of the three possible values - 0, if it stays in place, or \(\pm 1\), if it jumps.

**Observation 3.2** enables us to associate agent at location \(x\) to an integer value \(\lfloor x \rfloor\), its "integer" part. Now we need to decide which points are candidates to move, as more than one single point could potentially have the same integer coordinate. To settle the ambiguity we proceed as follows

**Definition 3.1.** Left/Right (Outer) Point (LP/RP) is a point with a minimum/maximum (integer) coordinate respectively.

![Figure 13: Possible constellation of few points.](image)

The original problem has exactly two Outer Points at any time \(t\), and the actual moving points are the current LP/RP. All the other points [still at time \(t\)] will be called further inner points.

We can define a new process:
• each point is associated with integer part of original coordinate, and could be thought as occupying one of the countably many bins labeled with the integers.

• the (arbitrary) points satisfying Definition 3.1 mimic the behavior of true left-most and right-most points in the original problem, i.e. LP in new problem jumps at the same direction as its left-most counterpart in original one.

• in case few points share same integer bin with LP/RP and the above decision moves such arbitrary point to next/previous bin respectively, we think of LP/RP ”pushing” one of its co-habitats in that direction (”further inside”) , and stay itself in the same bin. We call this a Push event.

![Figure 14: RP "pushing" inside point from bin 1 to bin 0.](image1)

The dynamic process move an outer point at each step when it is alone in its respective bin, but cases exist, when an outer point becomes an inner point and vice verse, and in this case we will have a Relabel event.

![Figure 15: Point relabeling after jump. LP "pushed" point to the right and RP jumped to the left.](image2)

Let \((X_k(t))_{t\geq 0}\) be a stochastic process describing a location of \(k^{th}\) point at time \(t\) in original (real coordinates) model and let \(Y_k(t) \triangleq \lfloor X_k(t) \rfloor\) for \(k \in \{1, 2, \ldots, n\}\) be a stochastic process tracking an ”integer” part of \(k^{th}\) point’s location, i.e. the coordinate in the integer model. Then, two next claims, proved in Appendix B ensure that switching to integer model and running it is equivalent to running the original model and switching to integer model only in the end.

**Claim 3.1.** Let \(i\) be an index of LP in tuple \((X_k(t))_{k=1}^n\) at time \(t\), then \(\forall k : Y_i(t) \leq Y_k(t)\).

Similar claim holds for RP. Note, that indexes \(k\) are fixed at the start, i.e. we labeled the points (see proof of Claim 3.1 in Appendix B). The claim help us track the outer points between the models, and the next lemma establishes the mathematical justification for casting conclusions back from integer to original model.
Lemma 1. Let $\chi : x \mapsto (\lfloor x \rfloor, \{x\})$ be a function splitting real numbers into pair of its integer and fractional parts, and $\pi_i : \mathbb{R}^n \to \mathbb{R}$ a standard projection of an $n$ dimensional vector onto its $i^{th}$ coordinate. Then, in case of Push/Relabel/"regular jump" event the next diagram is commutative

$$(X_k(t)) \xrightarrow{\text{time}} (X_k(t + 1))$$

$$\downarrow \pi_i \circ \chi$$

$$(Y_k(t)) \xrightarrow{\text{time}} (Y_k(t + 1))$$

Note, that $\pi_1 \circ \chi (\cdot) \equiv \lfloor \cdot \rfloor$ (see proof of Lemma 1 in Appendix B).

From here on, we will refer solely to the child process $Y$, i.e. two outer points will move left-right, when they encounter inner points they could "push" them in movement direction, at the same time staying at place, or they could eventually hit the wall, if all the points are at the same position.

In looking at the process, we may pin the coordinate frame to 0. But we can also transform the process $Y$, so that

<table>
<thead>
<tr>
<th><strong>Left-most inner point Assumption</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 will be attached to the <strong>left-most inner</strong> point.</td>
</tr>
</tbody>
</table>

We would further refer to the above as Assumption 3.

Observation 3.3. We note, that there are exactly two events, which lead to a shift of the time varying coordinate frame:

- **LP** is at 0 and pushes the **left-most** inner point to the right. The effect is double: LP "moves" from 0 to $-1$, every point except left-most inner point moves one step closer to 0. This event happens, only if LP shares location with exactly one inner point.

![Figure 16: LP "pushed" 0.](image)

- **RP** is at 0 and "pushes" some inner point to the left. Due to this, LP becomes 1 step closer to 0, all inner points, except left-most inner point, "pushed out" to 1 with RP.

![Figure 16: LP "pushed" 0.](image)
Recall, that in the case when all the points are at 0, moving LP, RP in opposite directions results in sending them to $-1, 1$ and all inner points stay at 0.
Chapter 4

Convergence of the Inner points

In this chapter we show that all inner points converge to an interval of length 1 for all possible values of $\varepsilon \leq \frac{1}{2}$, and also that for $\varepsilon < \frac{1}{2}$ convergence happens in a finite expected time less than $S_{c1}^{-2\varepsilon}$, where $S_c$ is the sum of all the initial coordinate differences from the left-most inner point.

Observation 4.1. If RP is at 0 necessarily all inner points are in the 0-bin.

Proof. Recall that by Assumption 3 there are no inner points to the left of the 0-bin. By Definition 3.1 there are no inner points to the right of RP either. Hence, if RP is at bin 0, then all inner points are there too.

We start by proving the central theorem of this chapter

Theorem 1 (Gathering of Inner Points). Let $Y_k(t)$ be as defined previously, with no labels attached to points (points have "addresses" or "indices" according to the bins they occupy). Let

$$L_I(t) = Y_{n-1}(t) - Y_2(t) + 1 = Y_{n-1}(t) + 1$$

- be the number of bins between the left-most and the right-most inner points, including the bins occupied by the left-most and the right-most inner points. Then for $\forall \varepsilon \in [0, \frac{1}{2}]$ there exists a finite time less than $T_{critical} < \infty$ WP 1 (i.e. with probability 1 or almost surely), such that $L_I(t)$ is monotonically non-increasing for $t \leq T_{critical}$ and $L_I(t) \leq 2$ for $\forall t \geq T_{critical}$.

To tackle the claim "there exists $T_{critical} < \infty$" we need to introduce the following technical lemma, which is proved in Appendix C

Lemma 2. Let $\xi_k = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } q = 1 - p \end{cases}$ be a sequence of i.i.d. random variables.

Let $S(t) = S(t-1) + \xi_t$ be a homogeneous random walk, starting at $S(0) = 1$. Define $T_x = \inf \{ S(t) = x \}$, then probability to visit 0 in a finite time is given by

$$r \triangleq P(T_0 < \infty | S(0) = 1) = \begin{cases} 1, & p \leq \frac{1}{2} \\ \frac{q}{p}, & \text{otherwise} \end{cases}$$

Remark 4.1. Before we proceed further, we want to establish a procedure for system state analysis. It is somehow cumbersome to analyze all 4 possible outcomes of LP/RP decisions at each step. We shall assume that LP and RP decide simultaneously, but then we will let them execute on those decisions in turn. Allowing RP to proceed, while deferring the movement of LP we analyze the intermediate state, then followed by LP movement, which brings the system to its final state (at that discrete time). Since the decisions are made prior to any changes made to the system, the actual order of motions, has no influence on the final state.
Remark 4.2. A multiset is a generalization of the concept of a set, that unlike a set, allows multiple instances of multiset’s elements. In the integer model the collection of point "addresses" falls into the definition of multiset.

Theorem 1 (Gathering of Inner Points). Let $Y_k(t)$ be as defined previously, with no labels attached to points (points have "addresses" or "indices" according to the bins they occupy). Assume, $Y_2(t) = 0$ (see Assumption 3). Let

$$\mathcal{L}_I(t) = Y_{n-1}(t) - Y_2(t) + 1 = Y_{n-1}(t) + 1$$

- be the number of bins between the left-most and the right-most inner points, including the bins occupied by the left-most and the right-most inner points. Then for $\forall \varepsilon \in [0, 1/2]$ there $\exists T_{\text{critical}} < \infty$ \textbf{WP 1} (with probability), such that $\mathcal{L}_I(t)$ is monotonically non-increasing for $t \leq T_{\text{critical}}$ and $\mathcal{L}_I(t) \leq 2$ for $\forall t \geq T_{\text{critical}}$.

Proof. The theorem is trivially true for $n \in \{1, 2, 3\}$ - just take $T_{\text{critical}} = 1$, so let $n > 3$. Observe, in general, that if multiset of inner points' coordinates does not change between discrete times $(t - 1)$ and $t$, then the value $\mathcal{L}_I(t)$ stays equal to $\mathcal{L}_I(t - 1)$ (from definition of $\mathcal{L}_I(t)$). Now, we should ask ourselves, whether it is possible to increase the value of $\mathcal{L}_I$ in one step? RP can only "push" inner points to the left, and LP to the right, otherwise they move outside - from the domain of the inner points. The action of RP could increase the number of bins occupied by inner points only in one case if RP "pushed" an inner point from $0$ to $-1$, which, due to Assumption 3 will become a new $0$. By the same assumption LP never can be to the right of $0$, and can increase the value of $\mathcal{L}_I$, by pushing some inner point to $1$. But, in order to increase the value of $\mathcal{L}_I$, no other inner point should be to the right of $0$ at the same point in time.

Hence, in both cases $\mathcal{L}_I(t) > \mathcal{L}_I(t - 1)$ implies immediately, that all the inner points at time $(t - 1)$ were at bin $0$. We conclude that next statements are equivalent

1. $\mathcal{L}_I(t) > \mathcal{L}_I(t - 1)$
2. $Y_2(t - 1) = Y_{n-1}(t - 1) = 0$
3. $\mathcal{L}_I(t - 1) = 1$

We have to show that there exists $T_{\text{critical}} < \infty$ with probability 1, such that $\mathcal{L}_I(T_{\text{critical}}) = 2$.

1. Define a new stochastic process $Z_i(t) \triangleq Y_n(t) - Y_n(0)$. This is a Standard biased Random Walk of RP. Define $T^i_{\text{push}} = \inf\{t : Z(t) = -1\}$ to be the first $(-1)$ Hitting Time. From Lemma 2 we have that for $\varepsilon < 1/2$ $P(T^i_{\text{push}} < \infty) = 1$, i.e. RP will eventually hit $(Y_n(0) - 1)$ almost surely.

2. Hitting on the $(-1)$ constitutes a "push" of an inner point, if co-located at $Y_n(0)$, or a move to $(Y_n(0) - 1)$ in the absence of the former.

3. As agents are memoryless, the dynamics of the stochastic process $Y(t - T^i_{\text{push}})$ is exactly the same as of $Y(t)$ due to spatial homogeneity, with an appropriate change of time. We conclude that for arbitrary starting location of the points, the time until "push" to the left is finite almost surely.

4. Since probability measure is \textit{countably subadditive}, we conclude that for any given $k \in \mathbb{N}$ the time until $k$ "pushes" to the left is finite almost surely. For this we define $B_i = \{\omega : T^i_{\text{push}}(\omega) < \infty\}$ - the set of all outcomes for which the $i^{th}$ "push" is made after finite time. Define $C_j = \bigcap_{i=1}^{j} B_i$ - all the outcomes for which the first $j$ "pushes" are made in finite time. We want to estimate $P(C_k)$. $P(B^C_i) = 0$ and $C^C_j \subseteq \bigcup_{i=1}^{j} B^C_i$, hence

$$P(C^C_j) \leq \sum_{i=1}^{j} P(B^C_i) = 0$$
and consequently $P(C_k) = 1$.

5. Each $(n-2)$ "pushes" to the left by RP, or to the right by LP, necessarily decreases the value of $\mathcal{L}_I$ by 1, if it is not yet 1. By careful accounting we note, that each $(n-2)$ "pushes" of any origin necessarily decrease the value by 1.

6. The initial value of $\mathcal{L}_I(0)$ is less than $\infty$. Each "push" takes a finite time almost surely, hence in a finite time $\mathcal{L}_I$ would decrease to 2 almost surely. This time we want to denote as $T_{\text{critical}}$.

**Corollary 1.** The gathered inner points will occupy only bins 0 and 1

**Proof.** Bin 0 is always occupied by definition, due to Assumption 3. From Theorem 1 it follows, that $\mathcal{L}_I(t) \leq 2$ for all $t \geq T_{\text{critical}}$. By recalling the meaning of the function $\mathcal{L}_I$ - the number of integer bins occupied by inner points at discrete time $t$, we conclude that if $\mathcal{L}_I(t) = 2$ for some $t > T_{\text{critical}}$, then bin 1 is also occupied at time $t$. □

The next logical step is to establish that time to "push" has a finite expectation for all $\varepsilon \in [0, 1/2)$ (see technical proof in Appendix C).

**Lemma 3.** Suppose RP moves according to movement rules. Let $Y_n(0) = x$ - be the initial coordinate of RP and define $T = \inf\{Y_n(t) = x-1\}$ to be the first time $t$ RP arrives to $(x-1)$.

We assume no inner point is at $Y_n(0)$. Suppose $E(T) < \infty$, then

$$E(T) = \frac{1}{1 - 2\varepsilon}$$

**Remark 4.3.** In the proof of Theorem 1 we defined $T_{\text{push}}^k$ - to be the time of the $k^{th}$ "push" to the left of RP. It follows that $T = T_{\text{push}}^1$.

Here we provide a simple proof, which does not require the deep martingale theory

**Proof.** We define $T_{x,x-1}$ to be the time it takes biased Random Walk to move from $x$ to $(x-1)$. Note, $T_{x,x-1} = T_{\text{push}}^1$ from the Remark 4.3

$$T_{x,x-1} = \begin{cases} 1 & \text{WP } 1 - \varepsilon \text{ Random Walk moves to the left} \\ 1 + T_{x+1,x} + T_{x,x-1} & \text{WP } \varepsilon, \text{ otherwise} \end{cases}$$

By writing out the expectation of $T_{x,x-1}$ we have

$$E(T_{x,x-1}) = 1 + \varepsilon \cdot (E(T_{x+1,x}) + E(T_{x,x-1}))$$

The Random Walk in question is shift-invariant, hence $E(T_{x+1,x}) = E(T_{x,x-1})$. Recall, that in the first proof we established $E(T_{x,x-1}) < \infty$, thus

$$E(T_{x,x-1}) = \frac{1}{1 - 2\varepsilon}$$

□

**Lemma 4.** If $Y_{n-1}(0) = Y_n(0)$, then the right-most inner point will be pushed once to the left in $\frac{1}{1 - 2\varepsilon}$ expected time.

**Proof.** Recall, that "push" is just the name of the event of RP jumping inside the inner points’ constellation, and the previous right-most inner point is relabeled to be a "new" RP. Hence, the claim directly follows from Lemma 3. □
Theorem 2. Denote by $S_C \triangleq \sum_{i=1}^{n} |Y_i(0)|$, where $Y_k(0)$ is the initial position of the $k^{th}$ point. Then the expected time until gathering of all the inner points at the same location (bin 0) is upper bounded by $S_C^1 - 2\epsilon$.

Proof. Due to Theorem 1 for all $t < T_{\text{critical}}$ LP’s action can only accelerate the gathering of the inner points, but never defer, we then evaluate the number of left “pushes” before $T_{\text{critical}}$ made by RP, ignoring any possible “help” from LP.

Denote by $Y_i = T_{\text{push}} - T_{\text{push} - 1}$ - time between $(i - 1)^{th}$ and $i^{th}$ pushes. By Lemma 4 $E(Y_i) = \frac{1}{1 - 2\epsilon}$. $k^{th}$ inner point should be pushed $Y_k(0)$ times to reach 0. RP itself should travel $Y_n(0) - 1$ steps towards 1, otherwise RP will be unable to push inner points to 0 from 1.

Denote by $\Upsilon \triangleq \sum_{i=1}^{S_C} Y_i$ - total number of left “pushes” by RP to concentrate all the inner points at 0. Then by linearity property of expectation we have

$$E(\Upsilon) = E(\sum_{i=1}^{S_C} Y_i) = \sum_{i=1}^{S_C} E(Y_i) = \frac{S_C}{1 - 2\epsilon} \quad (4.1)$$

Finally, we need to establish the upper bound, as in the statement of the theorem. Let $M$ - be the “meeting” location, i.e. $M = Y_2(T_{\text{critical}})$, which is a random variable. RP "pushed" at most $S_C$ times to the left, since it only needed to "push" the inner points, which were initially to the right of $M$, hence $S_C$, as defined, bounds the number of "pushes" made by RP before $T_{\text{critical}}$. As stated in the beginning of the proof, any "pushes" to the right by LP, made RP to "push" less times, before inner points converged.

Remark 4.4. Define $S(x) = \sum_{k=1}^{n} |x - Y_k(0)|$, then $S_C = S(Y_2(0))$. By negating the coordinates, LP and RP switch their roles, hence the upper bound on number of pushes could be further narrowed to

$$\min\{S_C, S(Y_{n-1}(0))\}$$

To estimate the probability that the convergence did not happened in say $m$ times the expected convergence time, we will use a standard technique of bounding such probability with a general Markov Inequality.

Lemma 5 (Markov’s Inequality). If $X$ is a nonnegative random variable and $a > 0$, then

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof. Let $I_{X \geq a}$ be an indicator random variable, that $X \geq a$, then it is obvious $aI_{X \geq a} \leq X$ or $I_{X \geq a} \leq \frac{X}{a}$. Now, it is trivial

$$E(I_{X \geq a}) \leq E(X/a) = \frac{E(X)}{a}$$

but

$$E(I_{X \geq a}) = 1 \cdot P(I_{X \geq a} = 1) = P(X \geq a)$$

□

We apply Markov Inequality to $\Upsilon$.

Theorem 3. Let $T_{\text{critical}}, S_C$ as defined above and $m > 0$, then

$$P(T_{\text{critical}} \leq \frac{mS_C}{1 - 2\epsilon}) \geq 1 - \frac{1}{m}$$
Proof. By Theorem 2 $\bar{Y}$ - is an upper bound on the number of "pushes" made by RP ($\bar{Y} \geq \text{T}_{\text{critical}}$), thus a nonnegative rv, hence from Lemma 5 for any $a > 0$

$$P(\bar{Y} \geq a) \leq \frac{\mathbb{E}(\bar{Y})}{a}$$

Take $a = m\mathbb{E}(\bar{Y})$, then

$$P(\bar{Y} \geq a) = P(\bar{Y} \geq m\mathbb{E}(\bar{Y})) \leq \frac{\mathbb{E}(\bar{Y})}{m} = \frac{1}{m}$$

it immediately follows

$$P(\bar{Y} \leq m\mathbb{E}(\bar{Y})) = 1 - P(\bar{Y} \geq m\mathbb{E}(\bar{Y}))$$

$$= 1 - \frac{1}{m}$$

$$P(\text{T}_{\text{critical}} \leq \frac{mS_{C}}{1 - 2\varepsilon}) \geq P(\bar{Y} \leq \frac{mS}{1 - 2\varepsilon}) \geq 1 - \frac{1}{m}$$

\[\square\]

4.1 A return to the original model

See Lemma 1 which justifies the validity of results deduced from the integer model to the original continuous one, to conclude that in original model we have the following result

**Corollary 2.** Inner points gather and stay in the interval of length 2.

**Proof.** Recall $Y_k(t)$ is an integer part of coordinate tracking process for point k. Due to Theorem 1 there exist $T_{\text{critical}} < \infty$, such that for any $t \geq T_{\text{critical}}$:

$$Y_{n-1}(t) - Y_2(t) \leq 1.$$ 

$$X_2(t) \geq Y_2(t)$$

$$X_{n-1}(t) \leq Y_{n-1}(t) + 1$$

hence, by putting above together

$$X_{n-1}(t) - X_2(t) \leq (Y_{n-1}(t) + 1) - Y_2(t) = (Y_{n-1}(t) - Y_2(t)) + 1 \leq 2$$

\[\square\]

What is more surprising is that we can actually bound the distance between inner points by 1.

**Theorem 4.** All the inner points gather to an interval of unit length $\mathbb{W} \mathbb{P} \ 1$ in finite time (and from that time, will forever occupy an interval shorter than a unit length).

**Proof.** Define

$$\mathcal{L}_{1t}(t) \triangleq X_{n-1}(t) - X_2(t)$$

- to be the distance between the left-most and the right-most inner points in the original model. Define $d = \min_{i \neq j} \{|X_i(0) - X_j(0)|\}$. $d$ is a constant and depends only on the initial constellation. It is obvious that in the process of jumps defined the fractional parts of the coordinates are independent of time (see Claim B.1 for the proof).

We want to show that there $\exists T_{\text{critical}}$ such that for $\forall t \geq T_{\text{critical}}$: $\mathcal{L}_{1t}(t) < 1$. Suppose that at some arbitrary time $t$, RP steps left and becomes one of the inner points. We analyze the intermediate state at time $t^+$ immediately after the jump following RP’s decision $\xi_t = -1$. 

17
1. If $L_{II}(t) > 1$, then $L_{II}(t^+) \leq L_{II}(t) - d$.

2. If $L_{II}(t) < 1$, then there are three possibilities

- It either became a non-left-most inner point, and we are in previous case, meaning $L_{II}(t^+) \leq L_{II}(t) - d < 1$.
- It became the left-most inner point, but $[X_2(t^+), X_{n-1}(t^+)] \subset [X_n(t) - 1, X_n(t)]$, hence $L_{II}(t^+) < 1$.
- It became a new LP, and we still have $[X_2(t^+), X_{n-1}(t^+)] \subset [X_n(t) - 1, X_n(t)]$ and $L_{II}(t^+) < 1$.

In any possibility $L_{II}(t^+) < 1$.

For the LP’s decision the things are symmetrical, as it executes own decision from $t^+$ if it was not an intermediate time, with one exception. In case RP’s decision left an intermediate state with a new LP, LP’s decision to step right makes it the new RP, but the inner points remain the same, hence $L_{II}(t^+1) = L_{II}(t) < 1$. We saw, that in case $L_{II}(t) > 1$ each step inside the group of the inner points decreases the value of the $L_{II}(t)$ by at least $d$. Hence in a finite number of such steps $L_{II}$ will decrease to the value less than 1, and, due to the above proof, will stay under 1 forever. As was shown in Lemma 4, the expected time between such "steps inside" is finite, hence in finite expected time $L_{II}$ will decrease to the value less than 1 for the first time. Denote this time as $T_{critical}^c$, obviously $T_{critical} \leq T_{critical}^c$.

See Figures 18 and 19 for the illustration of LP creating an intermediate state.

Figure 18: LP jumps inside

Figure 19: LP over-jumps inner points
Chapter 5

Bounding the outer point distributions

In this chapter we give an answer to the main question and show that in the long run the probability to find that system’s bounding interval exceeds $k$ in length is of the order $(\varepsilon^1 - \varepsilon^1)k$.

After inner points concentrated at $\{0, 1\}$ we need to establish the distribution of distances between outer points $LP$ and $RP$.

Observation 5.1. Inner points can be in $n - 2$ possible constellations. See Figure 20 for some possible constellations. Recall, that there is always at least one inner point at 0, due to Assumption 3.

So what happens to $LP$ and $RP$?

Let $\{W(t)\}_{t \geq 0}$ be a stochastic process tracking in time the coordinate of $RP$. Define, now, $\{X(t)\}_{t \geq 0}$ - $RP$’s coordinate in the case when $LP$ would be inactive and not pushing the left-most inner point. We start process $X$ at the same point, i.e. $X(0) = W(0)$. Clearly, $X(t) \geq W(t)$, since as mentioned earlier $LP$’s action is to cause all other points to get closer to 0. Let $\eta_t$ be the $t^{th}$ increment of $\{X(t)\}$. Note that $\eta_t$ depends on $X(t - 1)$ and the decision of $RP$. Define

$$\sigma_t = \begin{cases} 1, & \text{if } \eta_t = 1 \\ -1, & \text{otherwise} \end{cases}$$

and the process

$$Y(t) = \begin{cases} \max\{Y(t - 1) + \sigma_n, 1\}, & X(t - 1) \neq 0 \\ \max\{Y(t - 1) - 1, 1\}, & \text{WP } 1 - \varepsilon, \text{ if } X(t - 1) = 0 \\ Y(t - 1) + 1, & \text{WP } \varepsilon, \text{ if } X(t - 1) = 0 \end{cases}$$

Figure 20: Some possible inner point constellations. Number pair denotes a number of inner points in left and right bin accordingly.
We start \( \{Y(t)\} \) at \( Y(0) = \max\{X(0), 1\} \).

**Lemma 6.** \( \forall t : Y(t) \geq X(t) \)

*Proof.*

By definition, \( Y(0) \geq X(0) \). Now, we suppose that for all \( l \in \{1, \ldots, t - 1\} : Y(l) \geq X(l) \).

- If \( X(t - 1) > 0, \eta_t \neq 0 \), then \( \sigma_t = \eta_t \) and
  \[
  Y(t) = \max\{Y(t - 1) + \sigma_t, 1\} \geq Y(t - 1) + \sigma_m \geq Y(t - 1) + \eta_m \geq X(t - 1) + \eta_t = X(t)
  \]

- If \( X(t - 1) > 0, \eta_t = 0 \), then \( \sigma_t = -1 \). Note, that this situation is possible only for \( X(t - 1) = 1 \), i.e. RP pushes inner points from 1 to 0, then
  \[
  Y(t) = \max\{Y(t - 1) - 1, 1\} \geq 1 = X(t)
  \]

- If \( X(t - 1) = 0 \), then
  \[
  Y(t) \geq \max\{Y(t - 1) - 1, 1\} \geq 1 = X(t)
  \]

The last inequality follows from definition of \( \{X(t)\} \), if \( X(t - 1) = 0 \), then along RP’s decision, if it was to jump right or to push left-most inner point left, RP will find itself at 1.

\( \square \)

**Corollary 3.** \( \forall t : Y(t) \geq W(t) \).

**Lemma 7.** \( \{Y(t)\} \) is a Markov chain.

*Proof.*

We prove it by establishing transition probabilities \( P(Y(t) = y_t \mid Y(t - 1) = y_{t-1}) \) for all \( t \) and pairs of \( y_t \).

- If \( Y(t - 1) > 1 \)
  - In case \( X(t - 1) \neq 0 \) : then by definition
    \[
    \sigma_t = \begin{cases} 1, \varepsilon \\ -1, 1 - \varepsilon \end{cases}
    \]
    since RP moves right with \( \varepsilon \) probability, hence
    \[
    P(Y(t) = y_t \mid Y(t - 1) = y_{t-1}) = \begin{cases} \varepsilon, & \sigma_t = 1, y_t = y_{t-1} + 1 \\ 1 - \varepsilon, & \sigma_t = -1, y_t = y_{t-1} - 1 \\ 0, & \text{o.w.} \end{cases}
    \]
  - In case \( X(t - 1) = 0 \) : then by definition
    \[
    P(Y(t) = y_t \mid Y(t - 1) = y_{t-1}) = \begin{cases} \varepsilon, & y_t = y_{t-1} + 1 \\ 1 - \varepsilon, & y_t = y_{t-1} - 1 \\ 0, & \text{o.w.} \end{cases}
    \]

- If \( Y(t - 1) = 1 \), then all the previous points are still valid, except self-loop occurs, instead of left move, hence
  \[
  P(Y(t) = y_t \mid Y(t - 1) = 1) = \begin{cases} \varepsilon, & y_t = 2 \\ 1 - \varepsilon, & y_t = 1 \\ 0, & \text{o.w.} \end{cases}
  \]
Intuitively, \( Y \triangleq \{ Y(t) \}_{t \in \mathbb{N}} \) is a process, which tracks RP’s decisions on the line with partially absorbing wall at 1. By doing the same trick to the LP process (denote it as \( Z \)), we have two tracking processes, which are both bounded and biased Random Walks.

Let us look at \( Y \)'s underlying Markov chain

\[
\begin{array}{cccccccc}
\varepsilon & & & & & & \\
1 - \varepsilon & 2 & 3 & 4 & \cdots \\
\end{array}
\]

Figure 21: Simplified Markov chain associated with \( Y \)

This is an irreducible, ergodic Markov chain, hence by Theorem 10 of Appendix A if the stationary probabilities are found in such a Markov chain, then the solution is unique. Moreover the limiting probability distribution is the same distribution, since the chain is aperiodic [as a self-loop at state 1, ensures this is indeed the case].

To evaluate the stationary probability denoted by \( \pi(k) \) [i.e. the probability to find process \( Y \) at node \( k \)] we would cut the chain between state \( k \) and \( k - 1 \). Since probabilities are stationary, this process could be seen as evaluating the water flow inside and outside of the pool, while volume of water in pool stays constant. Means there is a 0 net flow through any cut in the chain. More commonly known as detailed balance equations, as \( Y \)'s Markov chain is also reversible.

We describe the equilibrium by the next equation

\[
\varepsilon \pi(k - 1) = (1 - \varepsilon) \pi(k)
\]  
(5.1)

or by isolating \( \pi(k) \)

\[
\pi(k) = \frac{\varepsilon}{1 - \varepsilon} \pi(k - 1)
\]

By successful substitutions we get

\[
\pi(k) = \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-1} \pi(1)
\]  
(5.2)

Since \( \pi \) is stationary probability of the chain, we also know

\[
\sum_{i=1}^{\infty} \pi(i) = 1
\]

We substitute Equation (5.2) into last equation to extract \( \pi(1) \)

\[
\sum_{i=1}^{\infty} \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{i-1} \pi(1) = \pi(1) \sum_{i=0}^{\infty} \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{i} = 1
\]

For \( \varepsilon < \frac{1}{2} \) this is a converging series, hence the solution could be found [and due to Theorem 10 unique].

\[
\Rightarrow \pi(1) = \frac{1}{\sum_{i=0}^{\infty} \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{i}} = 1 - \frac{\varepsilon}{1 - \varepsilon} = \frac{1 - 2\varepsilon}{1 - \varepsilon}
\]  
(5.3)

\( Y, Z \) are independent processes [\( Z \) is mirrored \( Y \), by putting "0" at right-most inner point and turning line direction we obtain the very same Markov chain], thus a basic question
\( P(\mathbf{Y}(t) + \mathbf{Z}(t) \geq k) \) is solved by a convolution, in our case, a discrete convolution of the distributions.

Since \( \mathbf{Y}(t), \mathbf{Z}(t) \) are independent, we break this probability into

\[
P(\mathbf{Y}(t) + \mathbf{Z}(t) \geq k)) = \sum_{i=1}^{k-2} P(\mathbf{Y}(t) = i)P(\mathbf{Z}(t) \geq k - i) + P(\mathbf{Y}(t) \geq k - 1) \tag{5.4}
\]

We estimate distribution tail

\[
P(\mathbf{Z}(t) \geq k) = \sum_{i=k}^{\infty} \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{i-1} \pi(1)
\]

but

\[
\sum_{i=k-1}^{\infty} \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{i} \pi(1) = \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-1} \pi(1) \frac{1}{\pi(1)} = \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-1}
\]

we then substitute into Equation (5.4)

\[
P(\mathbf{Y}(t) + \mathbf{Z}(t) \geq k) = \sum_{i=1}^{k-2} \left( \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{i-1} \pi(1) \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-i-1} \right) + \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-2}
\]

\[
= \sum_{i=1}^{k-2} \left( \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-2} \pi(1) \right) + \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-2}
\]

\[
= \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-2} (k-2)\pi(1) + 1
\]

and

\[
P(\mathbf{Y}(t) + \mathbf{Z}(t) < k) = 1 - \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-2} ((k-2)\pi(1) + 1) \approx 1 - k \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-2} \tag{5.5}
\]

This result answers the main question we were interested in - the probability to find all the points inside interval of the size \( k \) at time \( t \), as \( t \) goes to infinity (note stationary distributions of \( \mathbf{Y} \) and \( \mathbf{Z} \)).
Chapter 6

Convergence to stationary distribution of the outer points

We establish an upper bound on the convergence rate to stationary distribution for $\varepsilon < \frac{1}{3}$. We shall show it is strictly less than 1 by an amount we can estimate, hence the convergence is exponentially fast.

So we need to answer the following question: how fast does the $Y$ and $Z$ chains we considered in Chapter 5 converge to stationary distribution? We should note, that $Z$ starts at some point, which could be different from $Y$, but recall (Theorem 10) that limiting distribution is the same, for the two chains. To answer this question we use ideas developed in [Rosenthal 1996] and [Diaconis and Stroock 1991]. We want to establish an upper bound on the, so called, “second largest eigenvalue” of $P$ - the infinite transition matrix of the chains. Since underlying Markov chain of $Y$ is infinite, we use enlargements of finite subchains of $Y$ to bound convergence rate from above.

Let $X$ be a state space of $Y$’s Markov chain. Let $\pi$ be a stationary probability distribution on $X$, $\mu_0, \mu_k$ a probability distributions on $X$ after 0(initial) and $k$ steps respectively. We start by decomposing $X$ as $X = \cup_d X_d$, where each $X_d \subseteq X$ is finite, and $X_1 \subseteq X_2 \subseteq \ldots$. For $d$ large enough, we have $\mu_0(X_d) > 0$ and since $Y$’s chain is positive recurrent $\pi(X_d) > 0$. Let, then, $\pi_d$ be a probability measure on $X_d$ defined by $\pi_d(x) = \pi(x)/\pi(X_d)$ for $x \in X_d$, and similarly let $\mu_{0,d}(x) = \mu_0(x)/\mu_0(X_d)$ for $x \in X_d$. Further, define $P_d(x,y) = P(x,y)$ on $X_d$ for $x, y \in X_d$, $x \neq y$, and $P_d(x,x) = 1 - \sum_{x \neq y} P_d(x,y) = P(x,x) + P(x,X_d^C)$.

Those are general definitions. Our $Y$’s chain has a very specific structure, namely, it is a biased Random Walk with a partially reflective wall at 1. So, the decomposition is, in some way, natural: $X_d = \{1, 2, \ldots, d\}$. We can visualize $X_d$

![Figure 22: Finite Markov chain over $X_d$.](image)

Recall, that in original $Y$ $P(i,i+1) = \varepsilon$, while $P(1,1) = P(i+1,i) = 1 - \varepsilon$. $P_d(1,1)$ and $P_d(d,d)$ were calculated from the general definition above.
6.1 Markov Chain over $X_d$ and its properties

Denote by $P_d$ the transition matrix of a Markov Chain(MC)’s over $X_d$ (Figure 22).

$$P_d = \begin{pmatrix}
1 - \varepsilon & \varepsilon & 0 & \cdots & 0 \\
1 - \varepsilon & 0 & \varepsilon & \cdots & 0 \\
0 & 1 - \varepsilon & 0 & \varepsilon & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 1 - \varepsilon & 0 & \varepsilon \\
0 & \cdots & 0 & 0 & 1 - \varepsilon & \varepsilon
\end{pmatrix}$$

Claim 6.1. The MC over $X_d$ is ergodic [irreducible and aperiodic] for $\varepsilon \in (0, \frac{1}{2})$.

Proof.

i irreducible
The MC’s underlying graph is fully connected for $\varepsilon$ in setup. We assume directed edges are present only in case when transition probability is positive. Thus a path of length at most $d$ exists from any vertex to any other vertex with a non-zero transition probability. Hence the MC is a single recurrence class, and hence irreducible.

We should note, that due to finiteness if MC, this is positive recurrence class.

ii aperiodic
State 1 have a self-loop, hence its period is 1, and since the MC is a single recurrence class, every other state have the same period.

\[\square\]

Theorem 5. (Perron-Frobenius Theorem) Let $A = (a_{ij})$ be an $n \times n$ positive matrix: $a_{ij} > 0$ for $1 \leq i, j \leq n$. Then the following statements hold.

1. There is a positive real number $r$, called the Perron root or the Perron-Frobenius eigenvalue, such that $r$ is an eigenvalue of $A$ and any other eigenvalue $\lambda$ (possibly complex) is strictly smaller than $r$ in absolute value, $|\lambda| < r$. Thus, the spectral radius $\rho(A)$ is equal to $r$.

2. The Perron-Frobenius eigenvalue is simple: $r$ is a simple root of characteristic polynomial of $A$.

3. There exists an eigenvector $v = (v_1, \ldots, v_n)$ of $A$ with eigenvalue $r$ such that all components of $v$ are positive: $Av = rv$, $v_j > 0$ for $1 \leq j \leq n$. (Respectively, there exists a positive left eigenvector $w : w^T A = rw^T$, $w_j > 0$.)

The theorem could be extended to a non-negative matrices case, with obvious interchanges of positive for non-negative wordings. But for the case of irreducible matrices - it could be extended non-trivially. Namely, although the eigenvalues attaining the maximal absolute value may not be unique, the structure of maximal eigenvalues is under control: they have the form $e^{2\pi i k / h}$, where $h$ is an integer called the period of matrix, $r$ is a real strictly positive eigenvalue, and $k = 0, 1, \ldots, h - 1$. The eigenvector corresponding to $r$ has strictly positive components. Also all such eigenvalues are simple roots of characteristic polynomial.

Claim 6.2. The MC over $X_d$ has a unique stationary probability $\pi_d$.

Proof. Ergodicity ensures stationary distribution exists. $P_d$ is a stochastic matrix, hence have an eigenvalue of 1, and associated eigenvector $\tilde{\pi}_d : \tilde{\pi}_d P_d = \tilde{\pi}_d$. By Perron-Frobenius Theorem, 1’s algebraic multiplicity is 1. With additional constraint on the sum of eigenvector components [this is a probability distribution] we have a uniqueness. \[\square\]
Claim 6.3. The MC over $\mathcal{X}_d$ is reversible, given $\mathcal{Y}$ is reversible

Proof. Let write out generic left side for some $x, y \in \mathcal{X}_d$

$$\pi_d(x)P_d(x, y) = \frac{\pi(x)}{\pi(x_d)} P(x, y) = \frac{\pi(x)}{\pi(x_d)} \frac{\pi(x_d)P(x, y)}{\pi(x_d)}$$

reversibility

$$= \frac{\pi(y)}{\pi(x_d)} P(y, x) = \pi_d(y)P_d(y, x)$$

This is possible, due to the fact, that for any two different states in $\mathcal{X}_d$ the transition probability is the same as in original chain ($\mathcal{Y}$).

Since chain is reversible, probability distribution, which solves detailed balance equations is stationary distribution ($\hat{\pi}_d$), hence $\pi_d$ defined above, is one. To asses $\pi_d(x)$ for $x \in \mathcal{X}_d$ we calculate $\pi(\mathcal{X}_d)$

$$\pi(\mathcal{X}_d) = \sum_{i=1}^{d} \pi(i)$$

$$= 1 - \sum_{i=d+1}^{\infty} \pi(i)$$

$$= 1 - \sum_{i=d+1}^{\infty} \left(\frac{\varepsilon}{1-\varepsilon}\right)^{i-1} \pi(1)$$

$$= 1 - \left(\frac{\varepsilon}{1-\varepsilon}\right)^d \sum_{i=1}^{\infty} \left(\frac{\varepsilon}{1-\varepsilon}\right)^{i-1} \pi(1)$$

$$= 1 - \left(\frac{\varepsilon}{1-\varepsilon}\right)^d \cdot 1$$

$$= 1 - \left(\frac{\varepsilon}{1-\varepsilon}\right)^d$$

and

$$\pi_d(k) = \frac{\left(\frac{\varepsilon}{1-\varepsilon}\right)^{k-1} \pi(1)}{\pi(\mathcal{X}_d)}$$

hence

$$\pi_d(k) = \left(\frac{\varepsilon}{1-\varepsilon}\right)^{k-1} \pi_d(1)$$

This should not come as a surprise, as detailed balance equations are the same.

6.2 Transition probability matrix’s properties for reversible Markov chain.

Denote in general case

$$P_d = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,d} \\ p_{2,1} & p_{2,2} & \cdots & \cdots & p_{2,d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{d,1} & \cdots & \cdots & p_{d,d} \end{pmatrix}$$

Let

$$D = \begin{pmatrix} \sqrt{\pi(1)} & 0 \\ \sqrt{\pi(2)} & \cdots \\ \vdots & \ddots \\ 0 & \cdots & \sqrt{\pi(d)} \end{pmatrix}$$
We want to show $A = D\mathbb{P}_d D^{-1}$ is symmetric.

\[
A = D\mathbb{P}_d D^{-1} = \begin{pmatrix} \sqrt{\pi(1)} & 0 & \cdots & 0 \\ \sqrt{\pi(2)} & \sqrt{\pi(1)}p_{1,1} & \cdots & \sqrt{\pi(d)}p_{1,d} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sqrt{\pi(d)}p_{d,1} & \cdots & \sqrt{\pi(1)}p_{d,d} \end{pmatrix} \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,d} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ p_{d,1} & p_{d,2} & \cdots & p_{d,d} \end{pmatrix} D^{-1} = \begin{pmatrix} \sqrt{\pi(1)} & 0 & \cdots & 0 \\ \sqrt{\pi(2)} & \sqrt{\pi(1)}p_{1,1} & \cdots & \sqrt{\pi(d)}p_{1,d} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sqrt{\pi(d)}p_{d,1} & \cdots & \sqrt{\pi(1)}p_{d,d} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\pi(1)}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{\pi(2)}} & \frac{1}{\sqrt{\pi(1)}}p_{1,1} & \cdots & \frac{1}{\sqrt{\pi(d)}}p_{1,d} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{1}{\sqrt{\pi(d)}}p_{d,1} & \cdots & \frac{1}{\sqrt{\pi(1)}}p_{d,d} \end{pmatrix}
\]

Let's look at $(A)_{i,j}$:

\[
(A)_{i,j} = \begin{cases} \frac{\sqrt{\pi(i)}}{\sqrt{\pi(j)}} p_{i,j} & \text{if } (i,j) \text{ reversible} \\ \frac{\sqrt{\pi(j)}}{\sqrt{\pi(i)}} p_{i,j} & \text{otherwise} \end{cases} = (A)_{j,i}
\]

thus establishing symmetry.

**Claim 6.4.** Every eigenvalue of $\mathbb{P}_d$ is real.

**Proof.** $\mathbb{P}_d$ is similar to real symmetric matrix $A$, hence all its eigenvalues are real. \qed

### 6.3 Cheeger’s inequality

Let $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_m > -1$ be an eigenvalues of reversible finite $\mathbb{P}$ over $X$. Denote by $\lambda^* = \max\{|\lambda_2|, |\lambda_m|\}$. Let $\mathcal{M}$ be a set of all functions from $X$ to $\mathbb{C}$ and let $\mathbb{P}$ act on $\mathcal{M}$ by $(f\mathbb{P})(g) = \sum_{x} f(x)P(x,g)$. We can regard $\mu_k$ - temporal distributions of the chain as elements of $\mathcal{M}$ and $\mu_k = \mu_0 \mathbb{P}^k$. Define an inner product on $\mathcal{M}$ by $<f,g>_{L^2(1/\pi)} = \sum_{x\in X} f(x)g(x)/\pi(x)$

and $\|\cdot\|_{L^2(1/\pi)}$ induced by that inner product. In [Rosenthal, 1995] it is shown

**Claim 6.5.** Suppose reversible finite $\mathbb{P}$ satisfies $\lambda^* < 1$, then for an initial distribution $\mu_0$

\[
\|\mu_k - \pi\|_{TV} \leq \frac{1}{2} \|\mu_0 - \pi\|_{L^2(1/\pi)} (\lambda^*)^k
\]

where $\mu_k = \mu_0 \mathbb{P}^k$ and $\|\mu - \pi\|_{TV}$ is a total variation distance defined by

\[
\|\mu - \pi\|_{TV} = \sup_{\Lambda} |\mu(A) - \pi(A)| = \frac{1}{2} \sum_{x} |\mu(x) - \pi(x)|
\]

So we are interested in spectrum of $\mathbb{P}_d$, or more specifically in the closest to unit circle eigenvalue. General results on such eigenvalues are established in [Diaconis and Stroock, 1991], where Cheeger’s inequality is proved.
Let $P$ be a transition probability matrix of ergodic Markov chain with stationary distribution $\pi$. Assume chain is reversible with state space $X$. Define function $Q$ by

$$Q(x, y) = \pi(x)P(x, y) = \pi(y)P(y, x)$$

and for $A \subseteq X, B \subseteq X$

$$Q(A, B) = \sum_{x \in A, y \in B} Q(x, y)$$

Let $Y$ be some state subspace of $X$, then inequalities on the second largest eigenvalue of $P$ have been defined in terms of geometric quantity

$$h = \min_{\pi(Y) \leq 1/2} \frac{Q(Y \times Y^C)}{\pi(Y)} \quad (6.1)$$

We interchangeably use $h(Y)$ for a quantity $Q(Y \times Y^C)$ which is heuristically a measure of relative flow out of $Y$ when the chain is in stationarity. If $h(Y)$ is large for all $Y$, the chain should converge rapidly to $\pi$.

**Theorem 6. (Cheeger’s inequality)** Let $\lambda_2$ be a second largest eigenvalue of ergodic, reversible Markov chain with transition matrix $P$. Then

$$1 - 2h \leq \lambda_2 \leq 1 - \frac{h^2}{2}$$

with $h$ defined in Equation (6.1).

### 6.4 Applying Cheeger’s inequality

Denote by $c = \frac{\varepsilon}{1-\varepsilon}$. In our settings $c < 1$ for any $\varepsilon$. Recall, that speed of convergence is decided by $\lambda^* = \max\{|\lambda_2|, |\lambda_{m-1}|\}$. We need to establish a lower bound on $\lambda_{m-1}$ to get bounds on $\lambda^*$. There is a complicated way of doing this in the [Diaconis and Stroock 1991](#). Instead, we will apply a geometric technique to $P^2$. Obviously, $\lambda^*(P^2) = \lambda_2(P^2) \geq \lambda^*(P)^2$.

First, we establish the shape of Markov chain which corresponds to $P^2$. Then in the spirit of $P_d$ we will define $(P^2)_d$ s.

In $P^2$ there are edges that connect node $i$ to another node $j$, if there is a positive probability to move from $i$ to $j$ in two steps. Recall, that our original chain have a self-loop in 1, hence it is possible to get to/from 1 from/to 2 respectively in two steps. Other nodes could be achieved only from same-parity nodes at distance 2 by stepping twice in same direction. Figure 23 realizes this in Markov chain
Claim 6.6. \( \pi \) is stationary distribution of \( P^2 \).

Proof.

\[
\pi P^2 = (\pi P)P = \pi P = \pi
\]

To use Cheeger’s inequality in a new chain, we need to establish it is reversible.

Claim 6.7. Markov chain associated with \( P^2 \) is also reversible

Proof.

For each \( p^2(i, j) \) - probability to get from \( i \) to \( j \) in two steps we use a Chapman-Kolmogorov equality

\[
p^2(i, j) = \sum_{x \in \mathcal{X}} p(i, x)p(x, j)
\]

hence

\[
\pi(i)p^2(i, j) = \pi(i) \sum_{x \in \mathcal{X}} p(i, x)p(x, j)
\]

\[
= \sum_{x \in \mathcal{X}} \pi(i)p(i, x)p(x, j)
\]

\[
= \sum_{x \in \mathcal{X}} \pi(x)p(x, i)p(x, j)
\]

\[
= \sum_{x \in \mathcal{X}} \pi(j)p(j, x)p(x, i)
\]

\[
= \pi(j)p^2(j, i)
\]

and the chain is reversible.

Let \( \mathcal{X}_d \) be defined as above. See Figure 24 for visualization of Markov chain over \( \mathcal{X}_d \) for a two-step process.

Lemma 8. For \( \varepsilon < \frac{1}{4} \) subchain over \( \mathcal{X}_d \) which minimizes \( h \) is a continuous subchain from vertex 2 to the end of \( \mathcal{Y}^{2d} \) arm inclusive.

We notice that \( \pi_d(1) > \frac{1}{2} \), hence 1 is never a node in the candidate subchain, which minimizes \( h \). The next series of claims will help us to establish the proof.
Claim 6.8. Suppose $S_{k,j}$ is a continuous subchain in one of the arms, starting at node $k$ and ending at node $j$, such that $j$ is not the last node in the arm, then $h(S_{k,j}) = h(S_{k+2,j+2})$.

Proof. Imagine “moving” subchain to the right, with thick edges depicting probability flow out of respective subchain

$$h(S_{k,j}) = \frac{(1-\epsilon)^2 c^{k-1} + \epsilon^2 c^{j-1}}{c^{k-1} + \ldots + c^{j-1}} \cdot \frac{\epsilon^2}{c^2} = \frac{(1-\epsilon)^2 c^{k+1} + \epsilon^2 c^{j+1}}{c^{k+1} + \ldots + c^{j+1}} = h(S_{k+2,j+2})$$

Denote by $\text{Dist}(S,T)$ - distance between two subchains, a minimum distance between any pair of nodes of those subchains.

Claim 6.9. Let $S, T$ be two subchains over $X_d$, with $h(S) \leq h(T)$ and $\text{Dist}(S,T) \geq 2$. Then $h(S) \leq h(S \cup T)$.

Proof. Denote $h(S) = \frac{A}{B}$, $h(T) = \frac{C}{D}$, then as subchains have no neighboring nodes $h(S \cup T) = \frac{A+C}{B+D}$, i.e. flow out of both subchains accumulates, as does total weight. Then

$$\frac{A + C}{B + D} - \frac{A}{B} = \frac{AB + BC - AB - AD}{B(B + D)}$$

but

$$BC - AD = BD \left( \frac{C}{D} - \frac{A}{B} \right) \geq 0$$

hence $h(S) \leq h(S \cup T)$.

Claim 6.10. Let $S$ be a $S_{k,j} \cup S_{j+4,m}$, then $h(S_{k+2,m}) < h(S)$

Proof. We start by depicting the claim

$$h(S_{k,j}) = \frac{(1-\epsilon)^2 c^{k-1} + \epsilon^2 c^{j-1}}{c^{k-1} + \ldots + c^{j-1}} \cdot \frac{\epsilon^2}{c^2} = \frac{(1-\epsilon)^2 c^{k+1} + \epsilon^2 c^{j+1}}{c^{k+1} + \ldots + c^{j+1}} = h(S_{k+2,j+2})$$
By definition of $\pi_d$, $\pi_d(S_{k+2,j+2}) < \pi_d(S_{k+2,m})$, and
\[
Q(S_{k+2,j+2}, S_{k+2,m}) = \begin{cases} 
(1 - \varepsilon)^2 \pi_d(k + 2) + \varepsilon^2 \pi_d(j + 2), & m \text{ is last a node} \\
(1 - \varepsilon)^2 \pi_d(k + 2) + \varepsilon^2 \pi_d(m), & \text{o.w.} 
\end{cases}
\]
we combine both facts to establish $h(S_{k+2,m}) < h(S_{k+2,j+2})$. By Claim 6.9 $h(S_{k,j}) = h(S_{k+2,j+2})$, hence by Claim 6.3 $h(S_{k+2,m}) < h(S)$.

To summarize the last claim, in the quest for subchain with minimal $h$ it is worth to close "holes" in candidate subchain by right-shifting left ends. Now we can handle the main lemma in the chapter.

**Proof.**

Let look at general subchain $S$ - it is a collection of few continuous subchains. Claim 6.9 ensures, that each such subchain should be equally weighted by $h$ in order to be a candidate for subchain minimizing $h$, otherwise dropping $h$-heavy parts will create a lighter subchain. Claim 6.8 let us $h$-invariantly move subchains to the right, and with Claim 6.10 close gaps in subchain, including closing the gap with empty subchain! Hence we only need to take care of three groups of possible candidates to establish one, which minimizes $h$.

- subchain solely in the upper arm ("2"-arm)

![Figure 29: "2"-arm subchain.](image)

- subchain solely in the lower arm ("3"-arm)

![Figure 30: "3"-arm subchain.](image)
• subchain in both arms

![Diagram of a 3-arm subchain](image)

Figure 30: "3"-arm subchain.

We tackle first two points simultaneously.

1. We’ve established that minimum-weight subchain in one arm is the continuous tail, starting at some vertex \( k \). Denote it by \( S_k \) and by \( \hat{\delta} = \begin{cases} 1, & (d - k) \% 2 = 1 \\ 0, & \text{o.w.} \end{cases} \), i.e. indicator if \( k \) and \( d \) are in different arms.

\[
h(S_k) = \frac{Q(S_k, S_k^C)}{\pi_d(S_k)}
\]

\[
= \frac{(1-\varepsilon)^2 \pi_d(1)e^{k-1}}{\pi_d(1)(e^{k-1}+e^{k+1}+\ldots+e^{d-1})}
\]

\[
= \frac{1-2\varepsilon}{1-e^{d-k+1}}
\]

Recall,

\[(1-\varepsilon)^2(1-\varepsilon^2) = 1 - 2\varepsilon\]

2. By splitting into cases we have

(a) If \( k \) and \( d \) are in the same arm, then

\[
h(S_k) = \frac{1 - 2\varepsilon}{1 - e^{d-k+1}}
\]

(b) If \( k \) and \( d \) are not in the same arm, then

\[
h(S_k) = \frac{1 - 2\varepsilon}{1 - e^{d-k}}
\]

3. In any case \( h(S_k) \) monotonically increases with \( k \). Hence we need to compare two arms against each other to find the minimizing subchain.

(a) If \( 2 \) and \( d \) are in the same arm, then

\[
h(S_2) = \frac{1 - 2\varepsilon}{1 - e^{d-1}}
\]

while

\[
h(S_3) = \frac{1 - 2\varepsilon}{1 - e^{d-3}}
\]
(b) In the second case
\[ h(S_2) = \frac{1 - 2\varepsilon}{1 - cd^{-2}} \]
\[ h(S_3) = \frac{1 - 2\varepsilon}{1 - cd^{-2}} \]

In both cases \( S_2 \) had the least \( h \)-value.

4. As an alternative, an arm containing \( d \) node also achieves the minimum.

Note, that Claim 6.10’s proof - didn’t assume whole subchain is present only in single arm, hence we can conclude, that if subchain is split between both arms, then both subchains are continuous tails until the last node in the appropriate arm. Denote such subchain as \( T_{k,j} \), where \( k \) is the first node in the ”2”-arm, and \( j \) in ”3”-arm.

\[ h(T_{k,j}) = \frac{(1 - \varepsilon)^2 \pi_d(1)(e^{k-1} + e^{j-1})}{\pi_d(1)((e^{k-1} + \ldots + e^{d-1}) + (e^{j-1} + \ldots + e^{d-2}))} \]

\( d - 1 \) and \( d - 2 \) could, though exchange places, according to actual parity of \( d \). Let \( \hat{h} \) be kind of principal value of \( h \) and denote

\[ \hat{h}(S_k) = \frac{A}{B} = \frac{e^{k-1}}{e^{k-1} + \ldots + e^{d-1}} \]

and

\[ \hat{h}(S_j) = \frac{C}{D} = \frac{e^{j-1}}{e^{j-1} + \ldots + e^{d-2}} \]

(Once again assume WLOG the real placement of \( d - 1 \) and \( d - 2 \)). Due to Claim 6.9

\[ \hat{h}(T_{k,j}) = \frac{A + C}{B + D} = \frac{A}{B} = \hat{h}(S_k) \]

\[ \square \]

**Corollary 4.** \( h \geq \frac{1 - 2\varepsilon}{1 - cd^{-1}} \geq 1 - 2\varepsilon \)

So we’ve established an upper bound on \( \lambda^*(\mathbb{P}^2_d) \)

\[ \forall d : \lambda^*(\mathbb{P}^2_d) \leq \frac{1}{2} + 2\varepsilon - 2\varepsilon^2 \]

and we can extract an upper bound on \( \lambda^*(\mathbb{P}_d) \).

\[ |\lambda^*(\mathbb{P}_d)| \leq \sqrt{\lambda^*(\mathbb{P}^2_d)} \leq \sqrt{\frac{1}{2} + \varepsilon - \varepsilon^2} \quad (6.2) \]

Next task is to apply those uniform bounds to the original infinite chain. We start with technical claims, based on Proposition 2 in [Rosenthal 1996].

**Claim 6.11.** Let \( \mathbb{P}(\cdot, \cdot) \) be an irreducible Markov chain over \( \mathcal{X} \), reversible with respect to \( \pi(\cdot) \). Let \( \mathcal{X}, \pi_d, \mu_d, \mu_0, \mathbb{P}_d(\cdot, \cdot) \) be as above. Set \( \mu_k = \mu_0 \mathbb{P}^k \) and \( \mu_{k,d} = \mu_0 \mathbb{P}_d^k \). Then for each fixed \( x \in \mathcal{X} \) and \( k \geq 0 \), we have

- \[ \lim_{d \to \infty} \| \mu_{0,d} - \pi_d \|_{L^2(1/\pi_d)}^2 = \| \mu_0 - \pi \|_{L^2(1/\pi)}^2 \]
- \[ \lim_{d \to \infty} \pi_d(x) = \pi(x) \]
\[ \lim_{d \to \infty} \mu_{k,d}(x) = \mu_k(x) \]

**Proof.**

- We start by exploiting general properties of probability distributions \( \mu_0 \) - initial and \( \pi \) - stationary of some reversible chain \( P \).

\[
\|\mu_0 - \pi\|_{L^2(1/\pi)}^2 = \sum_{x \in X} \frac{(\mu_0(x) - \pi(x))^2}{\pi(x)} = \sum_{x \in X} \frac{\mu_0(x)^2 - 2\mu_0(x)\pi(x) + \pi(x)^2}{\pi(x)} = \sum_{x \in X} \frac{\mu_0(x)^2}{\pi(x)} - 1
\]

Hence,

\[
\|\mu_{0,d} - \pi_d\|_{L^2(1/\pi)}^2 = \sum_{x \in X_d} \frac{\mu_{0,d}(x)^2}{\pi_d(x)} - 1 \leq \frac{\pi(X_d)}{\mu(X_d)} \sum_{x \in X_d} \frac{\mu_0(x)^2}{\pi(x)} - 1 \to \sum_{x \in X} \frac{\mu_0(x)^2}{\pi(x)} - 1 = \|\mu_0 - \pi\|_{L^2(1/\pi)}^2
\]

- Recall the definition \( \pi_d(x) = \pi(x)/\pi(X_d) \). \( X_1 \subseteq X_2 \subseteq \ldots \subseteq X \) and \( \bigcup_d X_d = X \), hence \( \pi(X_d) \to \pi(X) \) and \( \pi_d(x) \to \pi(x) \).

- This claim is a bit more technically involved. Let write \( \mu_k(x) = P(A_d) + P(B_d) + P(C_d) \), where \( A_d \) is an event that the path of the original Markov chain ends in \( x \) (after \( k \) steps), but never leaves \( X_d \), \( B_d \) is an event it ends in \( x \), but leaves \( X_d \) at some point during the first \( k \) steps at least once, and \( C_d \) is an event it ends in \( x \), without leaving \( X_d \), with entering some self-loop at least once. On the other hand, since \( \mu_0,d(x) = \mu_0(x)/\mu_0(X_d) \) for \( x \in X_d \), we write \( \mu_{0,d}(x) = P(A_d)/\mu_0(X_d) + P(D_d) \), where \( D_d \) is an event that chain corresponding to \( P_d(\cdot, \cdot) \) over \( X_d \) ends in \( x \) and enters self-loop at least once. As previously, \( \mu_0(X_d) \to 1 \). Let \( \epsilon > 0 \) be given, denote by \( \rho = 1 - \frac{\epsilon}{\sqrt{1 - \epsilon}} \). Let \( S \subseteq X \) be such that \( \mu_0(S) \geq 1 - \rho \) and for each \( x \in S, S_x \subseteq X \), such that \( P(x, S_x) \geq 1 - \rho \) and \( |S_x| < \infty \). We note, \( S_x \) with above properties exists for each \( x \in S \), since \( P(x, X) = \sum_{y \in X} P(X, y) = 1 \), i.e. it is possible to pick a finite prefix in such series with requested properties. We choose \( d_0 \), such that \( S \cup \bigcup_x S_x \subseteq X_{d_0} \). The probability of starting or staying in \( X_{d_0} \) at any step is at least \( 1 - \rho \), hence \( P(B_{d_0}) \leq 1 - (1 - \rho)^{d+1} = \epsilon \), meaning \( P(B_{d_0}) \to 0 \). For the same reasons \( P(D_{d_0}) \to P(C_{d_0}) \), so that \( \mu_{k,d}(x) \to \mu_k(x) \).

\[ \square \]

**Claim 6.12.** In the same settings \( \lim_{d \to \infty} \|\mu_{k,d} - \pi_d\|_{TV} = \|\mu_k - \pi\|_{TV} \).

**Proof.**

Let \( \epsilon > 0 \) be given. We choose \( S \subseteq X \) with \( \pi(S) \geq 1 - \epsilon/8 \) and \( \mu_k(S) \geq 1 - \epsilon/8 \). Then we choose \( d_0 \), such that \( S \subseteq X_{d_0} \), and with \( |\mu_{k,d}(x) - \mu_k(x)| \leq \epsilon/8 |S| \) and \( |\pi_d(x) - \pi(x)| \leq \epsilon/8 |S| \) for all \( x \in S \) and for all \( d \geq d_0 \). Now,

\[ \pi_d(S) \geq \pi(S) \geq 1 - \epsilon/8 \Rightarrow \pi_d(S^C \mid X_d) \leq \frac{\epsilon}{8} \]

Since \( |\mu_{k,d}(x) - \mu_k(x)| \leq \epsilon/8 |S| \) it follows

\[ -\frac{\epsilon}{8 |S|} + \mu_k(x) \leq \mu_{k,d}(x) \leq \mu_k(x) \]
by summing over \( x \in S \)

\[
-\frac{\epsilon}{8} + \mu_k(S) \leq \mu_{k,d}(S)
\]

hence

\[
1 - \frac{\epsilon}{4} \leq \mu_{k,d}(S) \Rightarrow \mu_{k,d}(S^C|\pi_d) \leq \frac{\epsilon}{4}
\]

Then we are able to bound

\[
2 \|\mu_{k,d} - \pi_d\|_{TV} = \sum_{x \in X_d} |\mu_{k,d}(x) - \pi_d(x)|
\]

\[
= \sum_{x \in \mathcal{S}^c} |\mu_{k,d}(x) - \pi_d(x)| + \sum_{x \in \mathcal{S}} |\mu_{k,d}(x) - \pi_d(x)|
\]

\[
\leq \left( \frac{\epsilon}{2} + \frac{\epsilon}{8} \right) + \sum_{x \in \mathcal{S}} |\mu_{k,d}(x) - \pi_d(x)|
\]

\[
\leq \frac{\epsilon}{2} + \sum_{x \in \mathcal{S}} (\frac{\sqrt{\epsilon}}{\sqrt{\mathcal{S}}} + \frac{\sqrt{\epsilon}}{\sqrt{\mathcal{S}}} + |\mu_{k,d}(x) - \pi_d(x)|)
\]

\[
\leq \epsilon + \sum_{x \in X_d} |\mu_k(x) - \pi_d(x)|
\]

\[
= \epsilon + 2 \|\mu_k - \pi\|_{TV}
\]

It follows \( \limsup \|\mu_{k,d} - \pi_d\|_{TV} = \|\mu_k - \pi\|_{TV} \).

On the other hand

\[
\pi(S) \geq 1 - \frac{\epsilon}{8} \Rightarrow \pi(S^c) \leq \frac{\epsilon}{8}
\]

\[
\mu_k(S) \geq 1 - \frac{\epsilon}{8} \Rightarrow \mu_k(S^c) \leq \frac{\epsilon}{8}
\]

and

\[
2 \|\mu_k - \pi\|_{TV} = \sum_{x \in X} |\mu_k(x) - \pi(x)|
\]

\[
= \sum_{x \in X_d} |\mu_k(x) - \pi(x)| + \sum_{x \in \mathcal{S}} |\mu_k(x) - \pi(x)|
\]

\[
\leq \left( \frac{\epsilon}{2} + \frac{\epsilon}{8} \right) + \sum_{x \in \mathcal{S}} |\mu_k(x) - \pi(x)|
\]

\[
\leq \frac{\epsilon}{2} + \sum_{x \in \mathcal{S}} (\frac{\sqrt{\epsilon}}{\sqrt{|\mathcal{S}|}} + \frac{\sqrt{\epsilon}}{\sqrt{|\mathcal{S}|}} + |\mu_{k,d}(x) - \pi_d(x)|)
\]

\[
\leq \epsilon + \sum_{x \in X_d} |\mu_k(x) - \pi_d(x)|
\]

\[
= \epsilon + 2 \|\mu_k,d - \pi_d\|_{TV}
\]

It is obvious \( \liminf \|\mu_{k,d} - \pi_d\|_{TV} = \|\mu_k - \pi\|_{TV} \). The result follows. \( \square \)

**Lemma 9.** Under the above assumptions if \( \liminf \lambda^* (P_d) \leq \beta \), then

\[
\|\mu_k - \pi\|_{TV} \leq \frac{1}{2} \|\mu_0 - \pi\|_{L^2(1/\pi)} \beta^k
\]

hence the original (infinite) Markov Chain’s (MC) convergence rate is bounded by \( \sqrt{\left( \frac{1}{2} + \epsilon - 2\epsilon^2 \right)} \).

**Proof.**

It follows from [Equation (6.2)] and Claims [6.11] [6.12] and [6.5] Let \( \epsilon > 0 \) be given. Pick \( d_j \)-sequence, such that \( \lambda^* (P_{d_j}) \leq \beta + \epsilon \) for all \( j \). We have

\[
\forall j: \|\mu_{k,d_j} - \pi_{d_j}\|_{TV} \leq \frac{1}{2} \|\mu_{0,d_j} - \pi_{d_j}\|_{L^2(1/\pi_{d_j})} (\beta + \epsilon)^k
\]

34
by taking limits on both sides we have

$$\|\mu_k - \pi\|_{TV} \leq \frac{1}{2} \|\mu_0 - \pi\|_{L^2(1/\pi)} (\beta + \epsilon)^k$$

since $\epsilon$ is arbitrary this ultimately means

$$\|\mu_k - \pi\|_{TV} \leq \frac{1}{2} \|\mu_0 - \pi\|_{L^2(1/\pi)} \beta^k$$
Chapter 7

Simulations and results

We conducted a series of simulations to test empirically the theoretical results of the previous sections. Recall that the theoretical bound on the expected time to convergence was established in Equation (4.1), where $Y_i^0$ were the initial coordinates of agents, and is given by $\frac{1}{1-2\epsilon} \sum_{i=2}^{n} Y_i^0$.

Figure 32: Convergence Time vs. Length of Initial Interval.

Figure 33: Convergence Time vs. Number of Agents.
In the Figure 32 we present a functional relation between the length of initial interval and the time it takes to converge all the inner points to interval of the unit length. We ran on average 10 different iterations for all lengths between 10 to 1000 and for few different values of $n$. Convergence time vs. number of agents is presented in Figure 33. Note, that we previously established in Equation (4.1) linear dependency of convergence time from the sum of initial coordinates. As predicted by Lemma 4, convergence time is linear in both $n$ - number of agents, and the initial interval length. The latter comes from recalling the $k^{th}$ order statistics $X(k)$ of i.i.d. random variables $\{X_1, X_2, \ldots, X_n\}$, where $X_i \sim \text{Uni}(0, 1)$ is given by

$$X(k) \sim \text{Beta}(k, n - k + 1)$$

with

$$E(X(k)) = \frac{k}{n+1}$$

By adjusting to $n$ i.i.d. uniform random variables on $(a, b)$, we get

$$E(X(k)) = \frac{kb + (n - k + 1)a}{n+1}$$

We shall then adjust each statistic from $3^{rd}$ to $n^{th}$ by $E(X(2))$. Hence,

$$E(\sum_{k=2}^{n} X(k) - X(2)) = \sum_{k=3}^{n} \left[ \frac{kb + (n - k + 1)a}{n+1} - \frac{2b + (n - 1)a}{n+1} \right]$$

and after some algebra

$$E(\sum_{k=2}^{n} X(k) - X(2)) = \frac{n - 2}{2} \cdot \frac{n - 1}{n+1} \cdot (b - a) = O(n) \cdot (b - a)$$

Figure 34 depicts in more details the relation between length of initial interval for 100 agents, accompanied by predicted best/worst and average expected convergence times. Note, that average expected is hidden underneath the average actual simulation results, and best expected convergence time is the event of consecutive pushes, until convergence (or $\varepsilon = 0$). We’ve analyzed an upper bound in the worst case, but as previously remarked, the convergence time is roughly 4 times faster, than in the worst case, due to inner points converging towards the “middle” of the initial interval, as both outer points work simultaneously.

Figure 35 clearly shows that $\varepsilon = \frac{1}{2}$ is a breaking point, for which the convergence time is expected to be infinite. We should note, however, that the convergence of inner points is possible even for $\varepsilon > \frac{1}{2}$ (see Lemma 2 for the idea).
Figures 36 and 37 depict the simulation results for the distances between the outer points to the cluster of the inner points during the simulation. After the inner points converge to a 2-bin constellation, we have predicted the distribution tail to decrease exponentially fast and the pattern is indeed seen. As expected, the tail becomes long with increase in $\varepsilon$. Recall from Equation (5.5)

$$P(Y(t) + Z(t) \geq k) \approx k \left( \frac{\varepsilon}{1 - \varepsilon} \right)^{k-2}$$

And as $\varepsilon \to \frac{1}{2}$ the process becomes closer and closer to the standard Random Walk, where all the states are null-recurrent, i.e. we will see a 0 probability density at any given distance (state).
Chapter 8

Conclusions

We presented a 1-dimensional multi-agent convergence problem and a mathematical model, which made convergence claims easy to prove. We should note, that convergence happen to only part of the points [all except two], while two outside agents converged exponentially fast to some stationary distributed distance from the inner cluster of agents. This distance, at the bottom line, defines the diameter of agent distribution. And the probability to find outer agent at specific distance is itself decreases exponentially fast with the distance.

The main questions which remain unanswered around the current problem are: What is the distribution of the possible inner-points constellations? What are the convergence rates for $\varepsilon \geq 1/3$? To grasp the former question see the experiment results in Figures 38 and 39. As can be seen, the distribution of possible constellations is not uniform, with a notable state(s) $(n-2,0)$ (i.e. all the inner points in the same bin) frequency.

![Figure 38: Distribution of inner point constellations. $\varepsilon = 0.1$.](image)

![Figure 39: Distribution of inner point constellations. $\varepsilon = 0.4$.](image)

The additional complexity is added, by noticing, that constellations could be distinguished by parity of sum of all coordinates. In the Figures 38 and 39 states with parity 1 are depicted right after states with parity 0. Note that different parity is possible only for odd $n$s (imagine all the agents at some time $t_0$ are at 0, while at some other time $t_1$ at 1 to see the sum of coordinates does not have the same parity).

The next line of thought, is how can we extend these results to the original 2-dimensional problem. In this case more than 2 agents can jump at any given time step and the jumps could be of different step lengths?
Appendices
Appendix A

Markov Chains: Basic Theory

A.1 Definition and Examples

Definition A.1. A (discrete-time) Markov chain with (finite or countable) state space $\mathcal{X}$ is a sequence $X_0, X_1, \ldots$ of $\mathcal{X}$-valued random variables such that for all states $i, j, k, k_0, k_1, \ldots$ and all times $n = 0, 1, 2, \ldots$

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = k_n, \ldots) = p(i, j)$$ (A.1)

where $p(i, j)$ depends only on the states $i, j$, and not on the time $n$ or the previous states $k_n, k_{n-1}, \ldots$. The numbers $p(i, j)$ are called the transition probabilities of the chain.

Example A.1. The simple random walk on the integer lattice $\mathbb{Z}^d$ is the Markov chain whose transition probabilities are

$$p(x, x \pm e_i) = \frac{1}{2d} \quad \forall x \in \mathbb{Z}^d$$

where $e_1, e_2, \ldots, e_d$ are the standard unit vectors in $\mathbb{Z}^d$. In other words, the simple random walk moves, at each step, to a randomly chosen nearest neighbor.

Example A.2. The Ehrenfest urn model with $N$ balls is the Markov chain on the state space $\mathcal{X} = \{0, 1\}^N$ that evolves as follows: at each time $n = 1, 2, \ldots$ a random index $j \in [N]$ is chosen, and the $j^{th}$ coordinate of the last state is flipped. Thus, the transition probabilities are

$$p(x, y) = \begin{cases} \frac{1}{N} & \text{if the vectors } x, y \text{ differ at exactly one coordinate} \\ 0 & \text{otherwise} \end{cases}$$

The Ehrenfest model is a simple model of particle diffusion: Imagine a room with two compartments 0 and 1, and $N$ molecules distributed throughout the two compartments (customarily called urns). At each time, one of the molecules is chosen at random and moved from its current compartment to the other.

A.2 Chapman-Kolmogorov Equations and the Transition Probability Matrix

Assume henceforth that $\{X_n\}_{n \geq 0}$ is a discrete-time Markov chain on a state space $\mathcal{X}$ with transition probabilities $p(i, j)$. Define the transition probability matrix $\mathbb{P}$ of the chain to be

\[\text{[Based on the Lalley 2016]}\]
\( X \times X \) matrix with entries \( p(i, j) \), that is, the matrix whose \( i^{th} \) row consists of the transition probabilities \( p(i, j) \) for \( j \in X \):

\[
P = (p(i, j))_{i,j \in X}
\]  

(A.2)

If \( X \) has \( N \) elements, then \( P \) is an \( N \times N \) matrix, and if \( X \) is infinite, then \( P \) is an infinite by infinite matrix. Also, the \textit{row sums} of \( P \) must all be 1, by the law of total probabilities. A matrix with this property is called \textit{stochastic}.

\textbf{Definition A.2.} A \textit{nonnegative matrix} is a matrix with nonnegative entries. A \textit{stochastic} matrix is a square nonnegative matrix all of whose row sums are 1. A \textit{substochastic} matrix is a square nonnegative matrix all of whose row sums are \( \leq 1 \). A \textit{doubly stochastic matrix} is a stochastic matrix all of whose column sums are 1.

\textbf{Theorem 7.} The \( n \)-step transition probabilities \( p_n(i, j) \) are the entries of the \( n \)th power \( P^n \) of the matrix \( P \). Consequently, the \( n \)-step transition probabilities \( p_n(i, j) \) satisfy the Chapman-Kolmogorov equations

\[
p_{n+m}(i, j) = \sum_{k \in X} p_n(i, k)p_m(k, j)
\]

(A.3)

\textbf{Proof.} The Chapman-Kolmogorov equations are proved by double induction, first on \( n \), then on \( m \). The case \( n = 1, m = 1 \) follows directly from the definition of a Markov chain and the law of total probability (to get from \( i \) to \( j \) in two steps, the Markov chain has to go through some intermediate state \( k \)). The induction steps are proved equivalently. Once the equation is established, it follows, that \( n \)-step transition probabilities \( p_n(i, j) \) are the entries of \( P^n \), because \textbf{Equation (A.3)} is the rule of matrix multiplication.

Suppose now that the initial state \( X_0 \) is random, with a distribution \( \nu \), that is,

\[
P^\nu(X_0 = i) = \nu(i) \quad \forall i \in X
\]

Then by the Chapman-Kolmogorov equations and the law of total probability,

\[
P^\nu(X_n = j) = \sum_i \nu(i)p_n(i, j).
\]

equivalently, if the initial distribution is \( \nu^T \) (probability row vector on \( X \)) then the distribution after \( n \) steps is \( \nu^T P^n \). Notice that if there is a probability distribution \( \nu \) on \( X \) such that \( \nu^T = \nu^T P \), then \( \nu^T = \nu^T P^n \) for all \( n \geq 1 \). Consequently, if the Markov chain has such \( \nu \) as its initial distribution, then the marginal distribution of \( X_n \) will be \( \nu \) for all \( n \geq 1 \). For this reason, such a probability distribution is called \textit{stationary}.

\textbf{Definition A.3.} A probability distribution \( \pi \) on \( X \) is \textit{stationary} if

\[
\pi^T = \pi^T P
\]

(A.4)

\section*{A.3 Accessibility and Communicating Classes}

\textbf{Definition A.4.} A state \( j \) is said to be \textit{accessible} from the state \( i \) if there is a positive-probability path from \( i \) to \( j \), that is, if there is a finite sequence of states \( k_0, k_1, \ldots, k_m \) such that \( k_0 = i, k_m = j, \) and \( p(k_l, k_{l+1}) > 0 \) for each \( l = 0, 1, \ldots, m - 1 \). States \( i \) and \( j \) are said to communicate if each is accessible from the other. This relation is denoted by \( i \leftrightarrow j \).

\textbf{Fact A.1.} Communication is an equivalence relation. In particular, it is transitive: if \( i \) communicates with \( j \) and \( j \) communicates with \( k \) then \( i \) communicates with \( k \).
It follows that the state space $X$ is uniquely partitioned into communication classes. In general, if there is more than one communicating classes, then states in one communicating class $C_1$ may be accessible from states in another communicating class $C_2$; however no state in $C_2$ could be accessible from $C_1$.

**Definition A.5.** If there is only one communicating class (that is, every state is accessible from every other state) then the Markov chain (or its transition probability matrix) are said to be **irreducible**.

**Definition A.6.** The **period** of a state $i$ is the greatest common divisor of the set $\{n \in \mathbb{N} : p_n(i, i) > 0\}$. If every state has a period 1 then the Markov chain (or its transition probability matrix) is called **aperiodic**.

Note, if $i$ is not accessible from itself, then by convention we’ll define a GCD of an empty set to be $+\infty$.

**Example A.3.** Consider simple random walk on the integers. If at time 0 the walk starts in state $X_0 = 0$ then at any subsequent even time the state must be an even integer, and at any odd time the state must be an odd integer. Consequently, all states have period 2.

**Fact A.2.** If states $i,j$ communicate, then they have the same period. Thus, if the Markov chain is irreducible, then all states have the same period.

Hence, there is a simple test to check whether an irreducible Markov chain is aperiodic: if there is a state $i$ for which the 1-step transition probability $p(i, i) > 0$, then the chain is aperiodic.

**Fact A.3.** If the Markov chain has a stationary probability distribution $\pi$ for which $\pi(i) > 0$, and if states $i,j$ communicate, then $\pi(j) > 0$.

**Proof.** By [Definition A.3]

$$\pi(j) = \sum_{k \in X} \pi(k)p(k, j) \geq \pi(i)p(i, j) > 0$$

\[\square\]

## A.4 Finite State Markov Chains

### A.4.1 Irreducible Markov chains

If the state space is finite and all states communicate (that is, the Markov chain is irreducible) then in the long run, regardless of the initial condition, the Markov chain must settle into a steady state.

**Theorem 8** (Existence of Stationary Distributions Theorem). An irreducible Markov chain $X_n$ on a finite state space $X$ has a unique stationary distribution $\pi$. Furthermore, if the Markov chain is not only irreducible, but also aperiodic, then for any initial distribution $\nu$,

$$\lim_{n \to \infty} P^{\nu}(X_n = j) = \pi(j) \quad \forall j \in X$$  \hspace{1cm} (A.5)

### A.4.2 Standard simplex

**Definition A.7.** The subset $P$ of $\mathbb{R}^{|X|}$ gotten by intersecting the first orthant (the set of all vectors with nonnegative entries) with the hyperplane consisting of all vectors whose entries sum to 1 will be called **probability simplex** or **standard simplex**.

$P$ is closed and bounded, hence a compact subset of $\mathbb{R}^{|X|}$. We also define an action of $P$ on $P$ by $\nu^T \mapsto \nu^T P$. $P$ is a linear transformation, hence a continuous map from $P$ to itself.
A.4.3 The Krylov-Bogoliubov Argument

We can handle a first part of the Theorem 8 in a very simple way due to Krylov and Bogoliubov.

Fix an arbitrary probability vector $\nu \in P$, and consider the so-called Cesaro averages

$$\nu_n^T := \frac{1}{n} \sum_{k=1}^{n} \nu^T p^k$$

$\nu_n$ is a probability vector, hence $\nu_n \in P$. Then by Bolzano-Weierstrass the sequence $\lbrace \nu_n^T \rbrace$ has a convergent subsequence:

$$\lim_{k \to \infty} \nu_n^T = \pi^T$$

Claim A.1. The limit of any convergent subsequence of $\lbrace \nu_n^T \rbrace$ is a stationary distribution for $P$.

Proof. Denote the limit by $\pi$

$$\pi^T P = \lim_{k \to \infty} \nu_n^T P$$

$$= \lim_{k \to \infty} \frac{1}{n} \sum_{j=1}^{n} \nu_j^T p^j P$$

$$= \lim_{k \to \infty} \frac{1}{n} \sum_{j=2}^{n+1} \nu_j^T p^j$$

$$= \lim_{k \to \infty} \frac{1}{n} \left( \sum_{j=1}^{n} \nu_j^T p^j + \nu_j^T p^{n+1} - \nu_j^T p \right)$$

$$= \lim_{k \to \infty} \frac{1}{n} \sum_{j=1}^{n} \nu_j^T p^j$$

$$= \pi^T$$

The crucial part, is the existence of $\lim_{k \to \infty} \frac{1}{n_k} v = 0$ for any vector $v \in P$. 

A.4.4 Total Variation Metric

The most natural metric on the standard simplex $P$, turns, not the usual Euclidean, but a taxicab, called further the total variation distance, which is defined, as follows: for any two vectors $\mu, \nu \in P$,

$$d(\mu, \nu) = \|\mu - \nu\|_{TV} := 1/2 \sum_{i \in X} |\mu(i) - \nu(i)|$$

The factor $1/2$ is a long-standing convention (it ensures that the distance is never larger than 1). It could be shown

$$\|\mu - \nu\|_{TV} = \max_{A \subset X} \{\mu(A) - \nu(A)\}$$

Proposition A.1. Assume that entries of $P$ are all strictly positive. Then the mapping $\nu^T \mapsto \nu^T P$ is a strict contraction of the standard simplex $P$ relative to total variation distance, that is, there exists $0 < \alpha < 1$ such that for any two probability vectors $\mu, \nu$

$$\|\nu^T P - \mu^T P\|_{TV} \leq \alpha \|\nu^T - \mu^T\|_{TV}$$

Proof. Let $\varepsilon := 1/2 \min_{i,j} p(i,j)$, it follows $N\varepsilon < 1$, where $N = |X|$. Define $q(i,j) = \frac{p(i,j) - \varepsilon}{1 - N\varepsilon}$, and let $Q$ be the matrix with entries $q(i,j)$. Then $Q$ is a stochastic matrix, because for every state $i$

$$\sum_j q(i,j) = \frac{1}{1 - N\varepsilon} \sum_j p(i,j) - \frac{1}{1 - N\varepsilon} \sum_j \varepsilon = 1$$
Now consider the total variation distance between $\nu^T P$ and $\mu^T P$. By using the fact $\sum_i \nu(i) = 1$ and $\sum_i \mu(i) = 1$ we have

$$2 \|\nu^T P - \mu^T P\|_{TV} = \sum_j \left| (\nu^T P)_j - (\mu^T P)_j \right|$$

$$= \sum_j \left| \sum_i \nu(i) p(i,j) - \mu(i) p(i,j) \right|$$

$$= \sum_j \left| \sum_i (\nu(i) - \mu(i)) q(i,j) (1 - N \varepsilon) \right|$$

Let $\alpha := (1 - N \varepsilon)$. What left is

$$\sum_j \left| \sum_i (\nu(i) - \mu(i)) q(i,j) \right| \leq \sum_j \sum_i |\nu(i) - \mu(i)| q(i,j)$$

$$= \sum_j |\nu(i) - \mu(i)| \sum q(i,j)$$

$$= \sum_j |\nu(i) - \mu(i)|$$

$$= 2 \|\nu^T P - \mu^T P\|_{TV}$$

\[\square\]

### A.4.5 Proof of Theorem 8

For the proof we mention a famous Banach Fixed Point Theorem

**Theorem 9** (Fixed Point Theorem). Let $(X,d)$ be a non-empty complete metric space with a contraction mapping $T : X \to X$. Then $T$ admits a unique fixed point $x^*$ such that $T(x^*) = x^*$.

and note, that every compact metric space is complete.

**Lemma 10.** Let $P$ be the transition probability matrix of an irreducible, aperiodic, finite-state Markov chain. Then there is an integer $m$ such that for all $n \geq m$, the matrix $P^n$ has strictly positive entries.

**Proof.** It is enough to show, that $p_n(x,x) > 0$ for all large enough $n$s. Suppose that $p_n(x,x) > 0$ for all $n > m(x)$. Since the the Markov chain is irreducible, there exists $k = k(x,y)$ such that $p_k(x,y) > 0$, then from Chapman-Kolmogorov equations.

$$p_{k+n}(x,y) \geq p_n(x,x) p_k(x,y) > 0 \ \forall n \geq m(x)$$

Thus if $n \geq \max_{x} m(x) + \max_{x,y} k(x,y)$, then all the entries of $P^n$ will be positive.

It is not difficult to establish, that subset $A$ of $\mathbb{N}$ closed under addition, with GCD of its elements equal to 1, should contain all, but finitely many natural numbers. Defining

$$A_x := \{n \geq 1 : p_n(x,x) > 0\}$$

gives us such $A$ and closes the proof

Now, the $P$ itself should not be a strictly positive stochastic matrix, but by Lemma 10 there exists $m$, such that $Q := P^m$ is. As mentioned previously, $Q$ is also a stochastic matrix, which satisfies Proposition A.1, hence a contractive mapping on $P$. By the Banach Fixed Point Theorem 9 there exists a unique vector $\pi^T$, such that

$$\pi^T = \pi^T Q = \pi^T P^m$$

(A.6)
and such that for all $\nu \in \mathcal{P}$
\[
\lim_{n \to \infty} \nu^T Q^n = \lim_{n \to \infty} \nu^T P^m = \pi^T
\]
But this applies not only to $\nu^T$, but also to $\nu^T P, \nu^T P^2, \ldots$, since all those are probability vectors. Consequently, we have, for every $k = 0, 1, \ldots, m - 1$
\[
\lim_{n \to \infty} \nu^T P^m+k = \pi^T
\]
hence the whole sequence converge to the same limit
\[
\lim_{n \to \infty} \nu^T P^n = \pi^T \quad (A.7)
\]
It remains to show $\pi^T$ is a stationary distribution of $P$. Let $\mu^T := \pi^T P$. By Equation (A.6) it follows $\mu^T = \mu^T Q$. But $\pi^T$ is a unique stationary distribution of $Q$, hence $\mu = \pi$, and uniqueness follow from Equation (A.7).

## A.5 Stopping Times, Strong Markov Property

**Definition A.8.** Let $\{X_n\}$ be a Markov chain on finite or countable state space $\mathcal{X}$. A **stopping time** is a random variable $T$ with values in the set $\mathbb{N} \cup \{\infty\}$ such that for every $m \in \mathbb{N}$, the event $\{T = m\}$ is completely determined by the values $X_0, X_1, \ldots, X_m$.

**Example A.4.** The first passage time to state $x$ is a random variable $T_x$ whose value is the first time $n \geq 1$ that $X_n = x$, or $\infty$ if there is no such (finite) $n$. Clearly, $T_x$ is a stopping time

**Example A.5.** Let $L$ be the last time $n \leq 100$ that $X_n = x$ for some fixed state $x$. This isn’t (in general) a stopping time, because to determine whether $L = 93$, you would need to know not only the first 93 steps, but also all the steps from 94th till 100th.

**Proposition A.2** (Strong Markov Property). Let $T$ be a stopping time for Markov chain $\{X_n\}$. Then the Markov chain "regenerates" at time $T$, that is, the future $X_{T+1}, X_{T+2}, \ldots$ is conditionally independent of the past $X_0, X_1, \ldots, X_{T-1}$ given the value of $T$ and the state $X_T = x$ at time $T$. More precisely, for any $m < \infty$ and all states $x_0, x_1, \ldots, x_{n+m} \in \mathcal{X}$ such that $T = m$ is possible
\[
P(X_{T+i} = x_{m+i}, \forall 1 \leq i \leq n \mid T = m, X_i = x, \forall 0 \leq i \leq m) = \prod_{i=1}^{n} p(x_{m+i-1}, x_{m+i}) \quad (A.8)
\]

**Corollary 5.** State $x$ is recurrent if and only if the expected number of visits to $x$ is infinite
\[
E(N_x) = \sum_{n=0}^{\infty} p_n(x, x) = \infty \quad (A.9)
\]
where
\[
N_x = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = x\}}
\]

**Corollary 6.** Recurrence and transience are class properties. If $x$ is recurrent and $x$ communicates with $y$, then $y$ is recurrent.

Note, positive and null recurrence are also class properties. [Corollary 6] implies that in an irreducible Markov chain, all states have the same type (recurrent or transient). We’ll call irreducible Markov chain **recurrent or transient** according as its states are recurrent or transient (and similarly for positive and null recurrence).
A.6 Countable state space case

The existence and uniqueness of stationary distribution could be established also for a countable Markov chains. Some assumptions should be made and the next [Theorem 10] makes those precise.2

Let \( \tau_{ii} \) denote the return time to state \( i \) given \( X_0 = i \):

\[
\tau_{ii} = \inf\{ n \leq 1 : X_n = i | X_0 = i \}
\]

Let \( \pi_j \) denote the long run proportion of time that the chain spends in state \( j \):

\[
\pi_j = \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{1}_{\{X_m=j\}} \text{ WP 1}
\]

Theorem 10 (Existence and Uniqueness of stationary distribution). Let \( \{X_n\} \) be an irreducible positive-recurrent Markov chain, then a unique stationary distribution \( \pi \) exists and is given by

\[
\pi_j = \frac{1}{\mathbb{E}(\tau_{jj})} > 0, \text{ for all states } j \in \mathcal{X}
\]

If the chain is null recurrent or transient then the limits in Equation (A.10) are all 0 WP 1; no stationary distribution exists.

Proof. Following [Corollary 6] since the chain is irreducible, all the states belong to the same class, being either transient, positive recurrent or null recurrent.

First, we immediately obtain the transient case result since by definition, each fixed state \( i \) is then only visited a finite number of times; hence the limit in Equation (A.10) must be 0 WP 1.

Thus we need only consider the two recurrent cases.

First assume that \( X_0 = j \). Let \( t_0 = 0, t_1 = \tau_{jj}, t_2 = \min\{k > t_1 : X_k = j\} \) and in general \( t_{n+1} = \min\{k > t_n : X_k = j\}, n \geq 1 \). These \( t_n \) are the consecutive times at which the chain visits state \( j \). If we let \( Y_n = t_n - t_{n-1} \) (the interevent times) then we revisit state \( j \) for the \( n^{th} \) time at time \( t_n = Y_1 + \ldots + Y_n \). By the Strong Markov Property, the chain starts over again and is independent of the past every time it enters state \( j \). This means that the cycle lengths \( \{Y_n : n \geq 1\} \) form an iid sequence with common distribution the same as the first cycle length \( \tau_{jj} \). In particular, \( \mathbb{E}(Y_n) = \mathbb{E}(\tau_{jj}) \) for all \( n \geq 1 \).

Now observe that the number of revisits to state \( j \) is precisely \( n \) visits at time \( t_n = Y_1 + \ldots + Y_n \), and thus the long-run proportion of visits to state \( j \) per unit time can be computed as

\[
\pi_j = \lim_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} \mathbb{1}_{\{X_k=j\}} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Y_i = \frac{1}{\mathbb{E}(\tau_{jj})} \text{ WP 1},
\]

where the last equality follows from the Strong Law of Large Numbers, thus in the positive recurrent case, \( \pi_j > 0 \) for all \( j \in \mathcal{X} \), whereas in the null recurrent case, \( \pi_j = 0 \) for all \( j \in \mathcal{X} \).

If \( X_0 = i \neq j \), then we can first wait until the chain enters state \( j \) (which it will eventually, by recurrence), and then proceed with the above proof.

The uniqueness follows from the representation \( \pi_j = \frac{1}{\mathbb{E}(\tau_{jj})} \).

The more powerful would be the theorem establishing a necessary and sufficient condition for the existence of stationary distribution, and we quote here such a theorem without the proof (see Theorem 5.4 in [Gallager 2014]).

2Based on [Sigman 2009]
Theorem 11. Suppose \( \{X_n\} \) is an irreducible Markov chain (over \( \mathcal{X} = \{0, 1, \ldots\} \)) with transition matrix \( \mathbf{P} \). Then \( \{X_n\} \) is positive recurrent if and only if there exists a (non-negative, summing to 1) solution, \( \pi = (\pi_0, \pi_1, \ldots) \), to the set of linear equations \( \pi = \pi \mathbf{P} \), in which case \( \pi \) is precisely the unique stationary distribution for the Markov chain.
Appendix B

From real coordinates to integer model

We use this appendix to make formal claims about the validity of analyzing the integer model, instead of real numbers-based one.

Recall Observation 3.2 claimed, the fractional part of any point's location is unchanged throughout time. If we pair every real number $x$ with a tuple $(k, r)$, such that, $k \in \mathbb{N}, 0 \leq r < 1$ then

$$x = k + r$$

Let this function be $\chi : x \mapsto ([x], \{x\})$. By $\chi_2$ we denote an action of taking a fractional part of the number, formally a projection of $\chi$ to a second coordinate, i.e.

$$\chi_2 = \pi_2 \circ \chi$$

Claim B.1. Recall $(X_k(t))_{t \geq 0}$ is a stochastic process describing a location of $k^{th}$ point at time $t$. Then $\forall t : \chi_2(X_k(t)) = \chi_2(X_k(0))$.

Proof. Suppose for $t < \tau$ the claim is true.

$$X_k(\tau) = \begin{cases} 
X_k(\tau - 1) + 1, & \text{WP } \varepsilon \text{ to the right, } k \text{ is the rightmost point} \\
X_k(\tau - 1) - 1, & \text{WP } \varepsilon \text{ to the left, } k \text{ is the leftmost point} \\
X_k(\tau - 1) - 1, & \text{WP } 1 - \varepsilon \text{ to the left, } k \text{ is the rightmost point} \\
X_k(\tau - 1) + 1, & \text{WP } 1 - \varepsilon \text{ to the right, } k \text{ is the leftmost point} \\
X_k(\tau - 1), & \text{otherwise}
\end{cases}$$

Let $\chi(X_k(\tau - 1)) = (p, r)$ meaning $X_k(\tau - 1) = p + r$, hence

$$X_k(\tau) = \begin{cases} 
(p - 1) + r, & \text{moving left} \\
(p + 1) + r, & \text{moving right} \\
p + r, & \text{staying at same location}
\end{cases}$$

$$\Rightarrow \chi(X_k(\tau)) = ((p + a), r), \ a \in \{-1, 0, 1\}$$

$$\Rightarrow \chi_2(X_k(\tau)) = r,$$

Recall, that $\chi_2(X_k(\tau - 1)) = r$ and by induction hypothesis this is also $\chi_2(X_k(0))$.

Recall $Y_k(t) = \pi_1 \circ \chi(X_k(t))$ for $k \in \{1, 2, \ldots, n\}$ is a stochastic process tracking an "integer" part of $k^{th}$ point location. Next claim ensures outer point in the original process is not an inner point in process $Y$. 

53
Claim 3.1. Let $i$ be an index of LP(RP) in tuple $(X_k(t))_{k=1}^{n}$ at time $t$, then $\forall k : Y_i(t) \leq (\geq) Y_k(t)$.

Proof. Suppose point at index $j$ is the actual LP of tuple $(Y_k(t))_{k=1}^{n}$ and $Y_i(t) > Y_j(t)$. Then

$$X_j(t) = Y_j(t) + r_j(t) < Y_j(t) + 1 \leq Y_i(t) \leq Y_i(t) + r_i(t) = X_i(t)$$

contradicting the assumption, that $X_i(t) = \min_{1 \leq k \leq n} X_k(t)$.

For RP the proof is similar.

The opposite direction is not true, since few points could occupy same "integer" bin as could be seen in Figure 13.

Next lemma gives us mathematical justification to use process $(Y_k(t))$ instead of $(X_k(t))$.

Lemma 1. In case of "push"/"relabel"/"regular jump" event the next diagram is commutative.

\[
\begin{array}{c}
\begin{array}{c}
(X_k(t)) \\
\downarrow \pi_1 \circ \chi
\end{array} \\
\begin{array}{c}
(Y_k(t)) \\
\downarrow \pi_1 \circ \chi
\end{array}
\end{array}
\begin{array}{c}
\xrightarrow{\text{time}} \\
\xrightarrow{\text{time}}
\end{array}
\begin{array}{c}
(X_k(t+1)) \\
\downarrow \pi_1 \circ \chi
\end{array} \\
\begin{array}{c}
(Y_k(t+1)) \\
\downarrow \pi_1 \circ \chi
\end{array}
\]

Proof. As stated in Observation 3.1 only two points move in original process. So there are exactly 4 possible changes between $(X_k(t))$ and $(X_k(t+1))$, namely each of the end points (LP/RP) could move left or right. As $(Y_k(t))$ is a multi-set we need to ensure that same coordinates reoccur same number of times on both paths. For simplicity we assume WLOG that coordinates are ordered in the tuple, i.e. $X_1(t) < X_2(t) < \ldots X_n(t)$:

1. LP moved left [RP moved right]
   This is a simpler case, since by definition of LP
   $$X_1(t+1) = X_1(t) - 1 < X_1(t) \leq \min_{1 \leq k \leq n} X_k(t)$$
   this means $Y_1(t+1) = \pi_1 \circ \chi(X_1(t+1))$ is a unique occurrence in $(Y_k(t+1))$, namely $Y_1(t) - 1$. And this is also achieved in the child process, by moving LP to the left - its coordinate decreases by 1.

2. LP moved right [RP moved left]
   This situation could be split further into three cases
   (i) "relabel" event
   LP at time $t$ is not an LP after right jump. Collection with value $Y_1(t)$ decreased in size by 1, and collection with value $Y_1(t) + 1$ increased by one - on both paths.
   (ii) "push" event
   LP at time $t$ is not an LP after right jump. Collection with value $Y_1(t)$ decreased in size by 1, and collection with value $Y_1(t) + 1$ increased by one - on both paths. But, since, this is a "push" event collection of $Y_1(t)$ values had at least one inner point, which become LP, while LP become an inner point in $Y_1(t) + 1$.
   (iii) "regular" event LP left its position and $Y_1(t)$’s collection decreased in size by 1, while $Y_1(t) + 1$’s increased by one. By Claim 3.1 the diagram is commutative.

3. all points have the same "integer" coordinate at time $t$
   (i) opposite direction move
   LP/RP switches the roles and we use Claim 3.1 to establish commutativity. We can think of this as both end points "hit the wall" and moved with probability 1 back. See Figures [40] and [11] for illustration
   (ii) both endpoints moved in same direction
   WLOG LP moved left, so this case was already handled, while RP "pushed" inner point, which was also handled previously.

$\square$
Figure 40: LP/RP moved outside

Figure 41: LP/RP moved inside
Appendix C

Random Walks and Hitting Times

In this appendix we complement and rigorously prove claims regarding expected hitting time of a biased Random Walk on integers, we’ve used in our model.

We start by showing, that a biased Random Walk on integers started at 1, hits 0 with probability 1 for $p \in [0, \frac{1}{2}]$, and less than 1 for $p > \frac{1}{2}$, where $p$ - is the shift-to-the-right probability.

**Lemma 2.** Let $\xi_k = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } q = 1 - p \end{cases}$ be a sequence of i.i.d. random variables. Let $S(t) = S(t-1) + \xi_t$ be a homogeneous random walk, starting at $S(0) = 1$. Define $T_x = \inf\{S(t) = x\}$, then

$$r \triangleq P(T_0 < \infty \mid S(0) = 1) = \begin{cases} 1, & p \leq \frac{1}{2} \\ \frac{q}{p}, & \text{otherwise} \end{cases}$$

**Proof.** By the law of total probability

$$r = P(T_0 < \infty \mid S(0) = 1, S(1) = 0)P(S(1) = 0 \mid S(0) = 1) + P(T_0 < \infty \mid S(0) = 1, S(1) = 2)P(S(1) = 2 \mid S(0) = 1)$$

The process is shift invariant hence $P(T_1 < \infty \mid S(1) = 2) = P(T_0 < \infty \mid S(0) = 1)$. Additionally, to hit 0 starting from 2 we need to first hit 1 starting from 2, then hit 0 starting from 1. So,

$$r = 1 \cdot q + r^2 \cdot p$$

which could be solved for roots, giving

$$r_{1,2} = \frac{1 \pm \sqrt{1 - 4pq}}{2p}$$

and by substituting $q = 1 - p$

$$r_{1,2} = \frac{1 \pm (1 - 2p)}{2p}$$

yielding $r_1 = 1$ and $r_2 = \frac{1 - p}{p} = q/p$. For $p = 1/2$ roots coincide, while for $p < 1/2$ the value $r_2 > 1$, so it is not a valid probability.
For \( p > 1/2 \) let assume that \( r = 1 \). Hence, for almost all \( \omega \in \Omega \) there exist \( \tau_\omega \) such that \( S(\tau_\omega) = 0 \). If, \( S(\tau_\omega + 1) = -1 \), by the above argument \( P(T_0 < \infty \mid S(\tau_\omega + 1) = -1) = 1 \) (mirroring of the process). And using the law of total probability, we establish

\[
1 = P(T_0 < \infty \mid S(\tau_\omega) = 0) \\
= P(T_0 < \infty \mid S(\tau_\omega) = 0, S(\tau_\omega + 1) = -1) \cdot P(S(\tau_\omega + 1) = -1 \mid S(\tau_\omega) = 0) \\
+ P(T_0 < \infty \mid S(\tau_\omega) = 0, S(\tau_\omega + 1) = 1) \cdot P(S(\tau_\omega + 1) = 1 \mid S(\tau_\omega) = 0) \\
= 1 \cdot q + ? \cdot p
\]

Now, we are able to extract \(?\), giving us \( P(T_0 < \infty \mid S(\tau_\omega) = 0, S(\tau_\omega + 1) = 1) = 1 \). We conclude that \( S(t) = 0 \) for infinitely many times \( t \) with probability 1.

By the Strong Law of Large Numbers \( P(\lim_{t \to \infty} \frac{1}{t} S(t) = E(\xi_1)) = 1 \), which contradicts the immediate previous conclusion \( P(\liminf_{t \to \infty} \frac{1}{t} S(t) \leq 0) = 1 \), since \( E(\xi_1) = p - q > 0 \). Hence, for \( p > 1/2 \) the only possible value for \( r \) is \( q/p \).

**Lemma 3.** Suppose RP moves according to movement rules. Let \( Y_n(0) = x \) be the initial coordinate of RP and define \( T = \inf\{Y_n(t) = x - 1\} \) to be the first time \( t \) RP arrives to \( (x - 1) \).

We assume no **inner** point is at \( Y_n(0) \). Then for \( \varepsilon < 1/2 \) (i.e. \( E(T) < \infty \))

\[
E(T) = \frac{1}{1 - 2\varepsilon}
\]

**Proof.**

Without loss of generality assume that \( x = 0 \). Due to the definition RP’s motion is essentially a Random Walk biased towards the **inner** points and \( T \), as defined, is a \((-1) \) Hitting Time.

Let \( P \) be some path from 0 to \(-1\), with \(-1\) appearing for the first time as the last vertex on the path. It is impossible to arrive at \(-1\) from 0 in an even number of steps. And the last step on the path is from 0 to \(-1\). Hence, all the possible paths are of an odd length \( 2m + 1 \) for some \( m \).

To estimate the probability of such a path, that is realization of RP’s biased Random Walk, we need to calculate the probability of taking exactly \( m \) steps to the right in the first \( 2m \) steps, in a manner, that never moved left of the 0 in those \( 2m \) steps. This is a classical problem of generating all the possible ways to balance \( m \) pairs of parentheses. All such valid ways give rise to a valid path from 0 to 0 with the requested property, and all such paths are equiprobable.

The number of such paths are given by the \( C_m - \) the \( m^{th} \) Catalan Number. Assume \( H \) is an event, that \((-1)\) is hit, hence

\[
E(T \mid H) = \sum_{m=0}^{\infty} (2m + 1) \varepsilon^m (1 - \varepsilon)^{m+1} \cdot C_m
\]

\( C_m \sim \frac{4^m}{m^{\frac{3}{2}} \sqrt{\pi}} \cdot \) with \( \varepsilon < \frac{1}{2} \) it follows there exists \( c < 1 \), \( \varepsilon^m (1 - \varepsilon)^m \leq \left(\frac{1}{2}\right)^m c^m \). So a typical term in series is of \( O(m^{-\frac{3}{2}} c^m) \) and hence the series converges. Therefore we established, that the stopping time \( T \) has a finite expectation conditioned on \((-1)\) hit.

To proceed and extract \( E(T \mid H) \)'s value, we recall a famous Wald's Identity

**Theorem 12.** Let \( (X_m)_{m \in \mathbb{N}} \) be an infinite sequence of real-valued random variables and let \( N \) be a nonnegative integer-valued random variable.

**Assume that:**

1. \( (X_m)_{m \in \mathbb{N}} \) are all integrable (finite-mean) random variables
2. \( E(X_m 1_{N \geq m}) = E(X_m) P(N \geq m) \) for every natural \( m \)
3. \( \sum_{m=1}^{\infty} E(|X_m| 1_{N \geq m}) < \infty \)
4. \((X_m)_{m \in \mathbb{N}}\) all have the same expectation

5. \(E(N) < \infty\)

then 

\[
E(S_N) = E(\sum_{m=1}^{N} X_m) = E(X_1)E(N)
\]

Wald’s Identity

Let \(\xi_t\) be the decision of RP at time \(t\). By definition \(\xi_1, \xi_2, \ldots, \xi_m\) are i.i.d. with \(E(\xi_1) = 2\varepsilon - 1\). Define \(S_m = \sum_{t=1}^{m} \xi_t\).

Random Walk increments \(\xi_t\) and the Stopping time \(T\) fulfills the requirements of Theorem 12, hence

\[
E(S_T | H) = E(\sum_{t=1}^{T} \xi_t | H) = E(\xi_1)E(T | H)
\]

Left-hand side is known and is equal to exactly \(-1\) - we stop at \(x = -1\), hence

\[
E(T | H) = \frac{1}{1 - 2\varepsilon}
\]

(C.1)

We can write

\[
E(T) = E(T | H) \cdot P(H) + E(T | \mathcal{H} \cap \mathcal{C}) \cdot P(\mathcal{H} \cap \mathcal{C})
\]

Recall, \(P(H) = P(T < \infty)\), and due to Lemma 2 \(P(\mathcal{H} \cap \mathcal{C}) = 0\) for \(0 \leq \varepsilon \leq \frac{1}{2}\), i.e.

\[
E(T) = E(T | H)
\]
Bibliography


The triplet model of the location of the agents on elastic lines. The problem is identified in this case as a simple 1D problem. In this work, we analyze a simple 1D problem, which assumes that the movement of the agents is based on the existence of a neighbor on both sides of each agent.

Correctly, the agent identifies the direction of the other agents. When it is surrounded from both sides, it remains in its position, whereas the agents located at the end of the interval occupied by agents move.

Similarly to the original problem, only the agents located at the boundary of the system can move. In the 1D case, these are the two external agents from the right and left. The conclusion is that the distance between the two agents on the right and the two agents on the left is not monotonically increasing. These agents can change their "identities" and are not limited. We show that the size of the interval, which contains all the agents in a static state at time t, decreases as the length of the step decreases, in the expected time. Additionally, we show that the convergence probability is exponential and sets an upper limit to the time of convergence.
חוקי תנועת סוכן, הבתבסס על "האופק שב סוכן / מיפוי על ستוק". בלש
הדמויום הרד למוחוב מובחר, על החיה בלתי יעיל כל האפרכסר דוד התחלפ
מידע בעריך התכשיטות. במודלים קלאסיים, החששות של ווא, ואבד ביער, לבוח
ביוות-קושי, ניווט לפתרון מוקמי. לעמה את המסרכות מחובות סוכנים
ניידים, תקשורת חייה משמעית עד כדי ביל אפסיות כל文化节ים מוסים כלכל.
монтажים כללים. אתגרי לחקם את המודלים האָליאָל לקווגו על ידי ההנהלה
סוכנים ארבע יבולים ליגוד דר.

ש人造ה הוא קבץ סוכנים נוספים גורמים בתחלפ, בכי נַבְּשָּׁר מְשָׁמְעָה חישות בישות,}
מנבעת. בתאה דلازمאי התכשיטות בטבע, על חוקי תנועת הסוכנים לתחלפ
פroleum את הדיאדה של יוזonds מעכבים מדרכים. דורות שלפעים הסוכנים השונים
זָוֵיל הלָּטְנָכְּסֶתת בקְּנָוָה איוסו, ואַלְוָה את לווזיים מגזע. מסוכנים
מחומָּנָה של חישה: או מָנְיָס עייכים יולה רחרק שֶׁלשון. מסוכנים
אותרים.ecret נווה שמותיהם השופטיים חישה בור-顼וכים ולוסכנות גו
מעכרות קיוו מת DataGridViewCellStyle. או מנייסו שויו יולה אויר מידע על
הסיבובים, או סוכנים אתיר מברכה.

המודלTOTYPE ארבע בוחנים בעבדות ופושט מורא: מздоров של מעכרת
מרובים סוכנים על היל. הבתﭙה עֶשַׁת לארו עַנְייתים של בָּדָאִים
[Manor and Bruckstein, Manor and Bruckstein 2016] [Barel et al, 2016]
שקוור מָכּוֹמת אבדת בּניית התכשיטות של סוכנים על המיוּר. כו.
התיישנות והא הָּטָּוָנָה צוים, אֲלַמְּנֵוָטִום, שומואגוע לחרק את נמחה של
שוכנים ארואים, וזיוו לע פִיםְוַי הָּטָּוָנָה בָּזָבְּחָה. התהיה היא של סוכן
של תוחח ראו בּלָנָה מונבל, או התיישנותה השופטיים בָּדָאּים. ראה את
נכתות של סוכנים אתיר בחרי המשיוא שלחרי. במרחים מתמידים,
בּוּדָע שֶׁצרְפִיָה היא ויוה שהחלפין לאحتياו פֶּהאוּפָּט לשהיל שֶׁסרָּי
התקפות, המראים מרשים שֶׁסֶּרָּי אֵפוּטָר הָּטָּבְּחָה המדע המָעִים. ל soph.,
מסכה האָפָוּרָה היא שופניַס עוּךי הָּטָּוָנָה בָּדָאִים
לשתות. בתוכם, יש בוחר שישטו התכשיטות בָּדָאִים שֶׁלָּפְּיָה
כמַסָּר וַאָדוּת. בעֶיִים תָראוּשָיָל לָטָּבְּחָה וַיָּהוּ לָטָּבְּחָה מַמְּסָר מַמְּסָר.
וכל, בتراث על עַצְמְוָת פֶּרֶרָה פֶּהאוּפָּט עוּר על הבישת באָמוּתעה

תקציר

במשך אלפי שנים, השותאה האנושית נוכח며 מריבות ההנחות וההנחות השיתופית. התנסות בתנאים בולטים של נמלים, דבורים וטרמיטים, בלהקות של ציפורים ודגים, בנחילים של ארבה ובהתקבצויות של בני אדם.

למרות שבטבע התופעה היא מוכרת ושכיחה, ההתקדמים המדעיים של המאות האחרות שפכו מעט אור על המנגנונים המדויקים שבאמצעותם מיישמות הזהויות-facebook. אלא, בולע חיים בהתרימים ההפכים של קבוצות למעוננים של בני האדם השיתופיות שרוחם פtoISOString, את חספたく רחיצות ומתרחמות של תפרים לפלס, פلس ופלס.

ברובוטיקה, הפתרונות哈尔ים מתמטיקיים מתורמים כנוג אזור, טאוז וניקן של אוזר על隊 מרכרכה, אזור ואוביקטס ניווטים או ניידים בבסיסבת נציל-מרכחב. מספריים, ניט ולחלומו דע משמוד לתאלא לא-רורי-אילאיים, המותכניים במרכז עגור ממוכנות רובוטית במורבות יישומיות ובעלות כוח ייצור. אולא, לעבדות זו מציינה פсте של לזר. בתשואת התבנית, עם מועילות מגוון הסוכנים, נמצאת הקונוטציה עם התוכנות של התיאורית בשיתוף הרצויות, ואת חספたくさん מתרחמות של תפרים לפלס, פלס ופלס.

工作总结 (Swarm Robotics) של [Sahin, 2004], 살פי, התיאורית המרכבת ברמה המספקה את מופיעה התיאורית המספקת התיאורית של רבי חיות (חרקים) פשטות היחס. פיסטוס של סוכנים מתרחמות של תפרים לפלס, פלס ופלס. [Mohan & Ponnambalam, 2009] עליות כשמות, תורכים ממרכבות

לוגוUSH, המרכז הקהילה המודרנית בשיטת פסטרוס השבויים
הכללים, ברי היית, את עבורה התיאורית של קובץ הסוכנים ייחודיים
"הכותרת המובילה". טעדה בתורכים מיתות שיקופים ציונים של

המחקר עשה בהנחיית פרופ' אלפרד ברוקשטיין, בכיוון לימודי המחשב
אני מודה לטכניון על התמיכה הכספית הנדרשת להשלמתו.
התכנסות סוכנויות על ישר

חיבור על מחקר

לשם مليוי תחתיו של הדרישות לקבלת
הנאור מינטרו למדעי במדעי המחשב

דמיטרי רבינוביץ

הוגש לסנט הטכניון - מכון טכנולוגי לישראל
דצמבר 2017, חיפה, סגל ח""ק
התחכשות סוכנים על ישר

דmitry רבינוביץ',