Chapter 5

VERTEX-REINFORCED RANDOM WALK

ABSTRACT

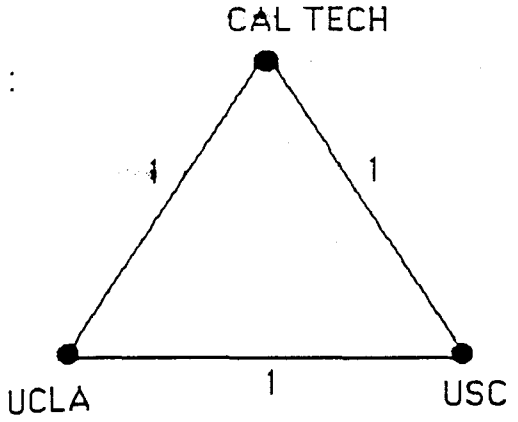
We define a discrete-time random process on a finite number of states. The transition probabilities at each time are influenced by the number of times each state has been visited, and by a fixed a priori likelihood matrix, $\mathbf{R}$. If $S_i(t)$ keeps track of the number of visits to state i up to time t, then we can form the fractional occupation vector, $\vec{v}(t)$, where $v_i(t) = S_i(t)/(\sum_{j=1}^{n} S_j(t))$. Under conditions on the non-degeneracy of $\mathbf{R}$, $\vec{v}(t)$ converges as $t \to \infty$ to a random vector in some finite set of points (theorem 5.8). The size of this finite set can be further reduced by ruling out some types of points it contains as possible places for $\vec{v}(t)$ to converge (theorems 5.11 and 5.12). The points of convergence that theorems 5.11 and 5.12 rule out are ones where the deterministic differential equation approximating the stochastic process has an unstable equilibrium.

5.1 Introduction

Imagine a visiting scholar newly arrived at UCLA. She drives back and forth between her offices at UCLA, USC and Cal Tech, always visiting different campuses on consecutive days. Her visits are initially random and unbiased, but she tends to visit a department more frequently the better she gets to know it. To model this, define a random walk on the vertices of a triangle. Initially all vertices are given weight 1, but the weight of a vertex increases by one every time the walk visits that vertex. The random walk begins at a specified vertex and moves by choosing among all neighboring vertices with probabilities proportional to the weights of the vertices.

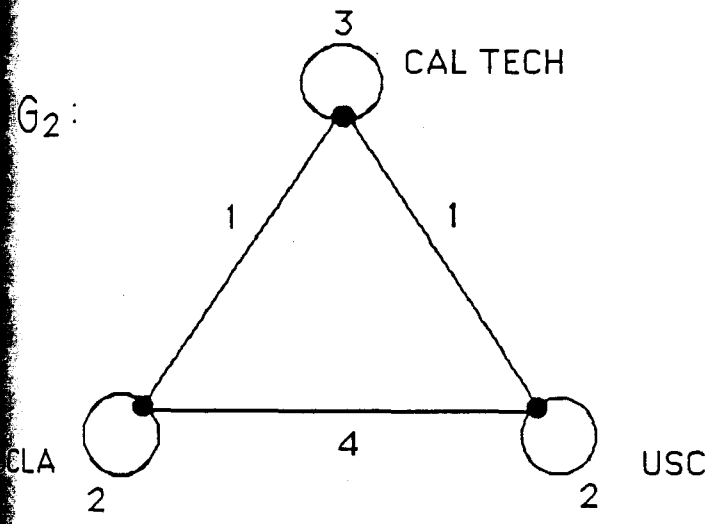
To make things more interesting, let's change the process so that visiting different campuses on different days is a preference but is not mandatory. Also, suppose that the visiting scholar hates to drive from LA to Pasedena or vice versa. To model the new process, add a loop at each vertex (i.e. an edge connecting the vertex to itself) and put weights on all the edges that represent an *a priori* likelihood of her driving along that edge. In figure 5, the graph G_2 has been assigned weights that might realistically model the visiting scholar process. The process proceeds similarly, with the vertices having initial weight 1 but increasing by 1 after each visit. This time, however the probability of a transition from a vertex to a given neighbor is proportional to the product of the *a priori* likelihood of a transition along the edge leading to that vertex and the weight of the vertex. Absent edges are taken to have likelihood zero. As an example, if the visitor begins at UCLA, the probabilities of her spending the first day at Cal Tech, UCLA and USC respectively are $1/7, 2/7$ and $4/7$. Say she spends the first day at USC; then the probabilities for the second day are respectively $1/9, 4/9$ and $4/9$. The likelihoods may be coded into a symmetric non-negative matrix which will always be denoted $\mathbf{R}$. The visiting scholar process will be called *vertex-reinforced random walk* (or VRRW) *with matrix* $\mathbf{R}$. The first graph in the figure 5 represents the simpler version of the visiting scholar problem, which will also be called *vertex-reinforced random walk* (or VRRW)

G_1:



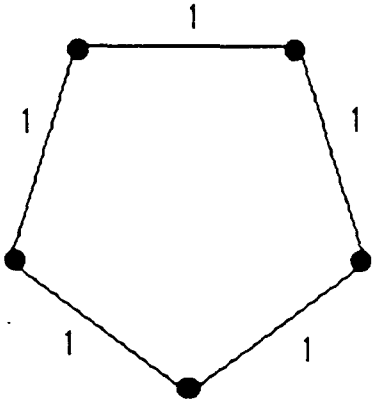
$$R = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

G_2:



$$R = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 2 \end{pmatrix}$$

G_3:



$$R = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

FIGURE 5

on the triangle. In general, VRRW on a graph means VRRW with the incidence matrix for that graph.

More formally, define the process, vertex-reinforced random walk, to be a sequence $d \times d$ matrix with non-negative entries. Think of $\mathbf{R}$ as the matrix of conductivity between pairs of states. Let $Y_d = 1$ and let $\vec{S}(n)$ be a vector with coordinates $S_i(n) = 1$ plus the total number of times that Y_j has been in each state for time between $d+1$ and n, so for example $\vec{S}(d+1) = (1, 1, \cdots, 2)$ whenever $Y_{d+1} = d$. Formally,

$$S_i(n) = 1 + \sum_{j=d+1}^{n} \delta_{i,Y_j}. \tag{1}$$

Let the transition probabilities for Y be defined by $prob(Y_{n+1} = r \,|\, Y_n = s, \vec{S}(n)) = \mathbf{R}_{sr}\vec{S}(r)/\sum_\alpha \mathbf{R}_{s\alpha}\vec{S}(\alpha)$. Formally the process is defined on a measure space Ω; we will use ω to denote points in Ω, the few times the need arises.

Remark: The initial condition $\vec{S}(d) = (1, \ldots, 1)$ is completely immaterial. All the theorems in this chapter describe which kinds of limiting behaviour are possible and which are not. The are equally valid for any nonzero starting weights and the particular choice of all 1's is just for convenience.

Remark: The symmetry of $\mathbf{R}$ is important for all the calculations involving eigenvalues and fixed points. I believe that the main theorems are true as well when $\mathbf{R}$ is not symmetric but I have been unable to prove them since they rely on proposition 5.4 which involves a calculation of a fixed point.

Define

$$\vec{v}(n) = \vec{S}(n)/n \tag{2}$$

so that $\vec{v}$ is just $\vec{S}$ normalized to have its components sum to 1. Often we treat $\vec{v}$ as the object of study, rather than Y or $\vec{S}$, so the sample paths of the process are viewed

as sequences of points in the $n-1$-simplex $\triangle \subseteq \mathbf{R}^n$. If some diagonal entries of $\mathbf{R}$ are 0, then the coordinates of $\vec{v}$ for those states can never be more than $1/2$. Nevertheless, most of the theorems in this chapter are true for vectors ranging over the entire simplex $\triangle$. One important quantity has a special notation because it is used so often.

Definition 5.1 *Let* $N_i(\vec{v}) = \sum_r \mathbf{R}_{ir} v_r$. *Let* $N_i(n)$ *(or just* N_i *when the* n *is understood) denote* $N_i(\vec{v}(n))$.

In the case where $\mathbf{R}$ is an incidence matrix, $N_i(\vec{v})$ is just the sum of values of $\vec{v}$ over all neighbors of i. At the risk of abusing notation $\vec{v}$ will be used for a generic point in $\triangle$, while $\vec{v}(n)$ denotes the sample occupation vectors of the vertex-reinforced random walk.

Vertex-reinforced random walks behave quite differently from edge-reinforced random walks, even though the set-up is apparently similar. Recall that for edge-reinforced random walks $\vec{v}(n)$ converges as $n \to \infty$ to a random vector whose distribution over the simplex is absolutely continuous with a density that can be explicitly described. In the case of Vertex reinforcement, for most matrices $\mathbf{R}$, it is still true that $\vec{v}(n)$ converges to a random vector, but the distribution of the random vector is far from being absolutely continuous. In general, the measure is supported on a discrete set of points (theorem 5.8 and corollary 5.9).

Most of what follows can be intuitively understood by seeing what is special about this discrete set of points. Here is the heuristic explanation. Suppose n is large. Then the transition probabilities vary slowly, and the process looks like a Markov chain over a span of L steps, for $L \ll n$. Furthermore, if $1 \ll L \ll n$ then no matter what state the process is in at time n, the numbers of visits to each state over the next L steps is approximately L times the stationary distribution, call it π, for the Markov chain. Later in this chapter, in the proofs of lemma 5.7 and proposition 5.21, versions of the

approximation

$$\mathrm{E}(\vec{v}(n+L) - \vec{v}(n) | \vec{v}(n)) \approx \frac{L}{n}(\pi(\vec{v}(n)) - \vec{v}(n)).$$

In general, the stationary distribution $\pi(\vec{v}(n))$ is not the same as $\vec{v}(n)$, so $\vec{v}$ begins to change. In fact the process behaves like a multicolor nonlinear urn, or a stochastic approximation with correction term in the direction of $\pi(\vec{v}(n)) - \vec{v}(n)$. For some choices of $\vec{v}$ the stationary distribution $\pi(\vec{v})$ is just $\vec{v}$ again. This is a necessary condition for convergence of $\vec{v}(n)$ to the point $\vec{v}_0$ with non-zero probability. (This is where vertex-reinforcement differs from edge-reinforcement: in edge-reinforcement, the stationary distribution is always given by the current value of $\vec{v}(n)$, so there is no bias in the direction of drift of this vector. The set of points where there is no drift in the vertex-reinforced case is typically discrete.)

Making an analogy with differential equations as in [NH], the motion of $\vec{v}(n)$ may be thought of as an autonomous flow, $\vec{v}(t)$ following a vector field $\vec{F}$ so that $d/dt(\vec{v}(t)) = \vec{F}(\vec{v}(t))$. Here, $\vec{F}(\vec{v})$ is the heuristic analogue of the expected change in the VRRW form time n to time $n + L$, which depends only on $\vec{v}(n)$ after a time change. The previous paragraph identifies the direction of $\vec{F}(\vec{v})$, as being a vector pointing from $\vec{v}$ toward $\pi(\vec{v})$. This flow has a discrete set of critical points $\vec{v}$ for which $\pi(\vec{v}) = \vec{v}$. Some of these equilibria are unstable under small perturbations; it turns out that $\vec{v}(n)$ cannot converge to these unstable points (theorems 5.11 and 5.12). Thus the necessary condition $\pi(\vec{v}) = \vec{v}$ is not sufficient for convergence to $\vec{v}$ with non-zero probability. These theorems are almost a corollary of theorem 3.8 on non-convergence of generalized urn processes to unstable equilibrium points of the associated flow. The same method of proof is used, and the arguments are long only because of technical difficulties in stting up VRRW as a generalized urn process.

To illustrate a few of the thigs that can happen, consider the three cases shown in figure 5. When a VRRW is run on the graph G_1, the vector $\vec{v}(n)$ always converges to $(1/3, 1/3, 1/3)$, so the visiting scholar pends asymptotically equal time at each of the

three math departments. When the second VRRW is run, it turns out that $\vec{v}(n)$ must converge to one of the two points $(1,0,0)$ or $(0,1/2,1/2)$. In other words the visitor ends up either spending asymptotically all her time at Cal Tech or asymptotically ignoring Cal Tech and dividing her time equally between UCLA and USC. It is random which of these actually happens.

The VRRW for G_3 is interesting because the theorems of this chapter do not cover it. In fact while the theorems apply to "almost all" symmetric non-negative matrices $\mathbf{R}$, they often fail in the case that are naturally interesting, since the two conditions they require are that $\mathbf{R}$ not have entries of the diagonal that are zero and that its eigenvalues be linearly independent over the rationals, and in particular distinct. (The eigenvalue condition may not be difficult to remove eventually, but the other condition is difficult to get around.) While the set of matrices that fail these conditions is small – the union of several closed manifolds in matrix space – it contains all the examples with a lot of symmetry, which are the ones that are natural to look at. In these cases, we must be satisfied with information such as "if VRRW is run with a matrix given by any small generic perturbation of $\mathbf{R}$, then . . ." This is in fact what led me to study VRRW with general matrices instead of sticking to VRRW on graphs: the general case is solvable, though the particular case is not!

Although I cannot prove it, I will state what happens to VRRW on G_3. The walk asymptotically ignores two adjacent sites (which two is random) and visits the middle of the remaining three on alternate moves, spending the remaining half the time at the last two vertices in a random limiting proportion. So $\vec{v}(n) \to (x, 1/2, 1/2 - x, 0, 0)$ for some x, or it converges to one of the other four types of vector gotten from this type by cyclically permuting the indices. The foregoing examples, along with a few others are worked out in section 5.6.

5.2 A Liapunov function for the motion of $\vec{v}(t)$

The method of establishing the convergence of $\vec{v}(n)$ is to show that the distance between $\vec{v}(n)$ and a set of critical points converges to 0. When this set is discrete, the convergence of $\vec{v}(n)$ follows. To do this, we find a function $H : \Delta \to \mathbf{R}^+$, which always increases when the vector $\vec{v}(t)$ changes infinitesimally in the direction of $\mathbf{E}(\vec{v}(n+L) \mid \vec{v}(n))$ for an appropriate constant, L. In the analogy with a differential equation, this is a Liapunov function for the flow

$$\frac{d}{dt}\vec{v}(t) \;=\; \pi(\vec{v}(t)) - \vec{v}(t) \;\approx\; \frac{t}{L}\mathbf{E}(\vec{v}(t+L) - \vec{v}(t)|\vec{v}(t)). \tag{3}$$

Lemma 5.7 identifies this function. The set that $\vec{v}(n)$ must approach is the set of critical points for the potential function and the main work is showing that the actual (stochastic) process $\vec{v}$ behaves like the deterministic flow $\vec{v}(t)$. For those unfamiliar with the method of Liapunov functions, here's how the argument goes. In the continuous time flow (3), $H(\vec{v}(t))$ is bounded and cannot increase forever except by increasing more and more slowly. The rate of increase of H is just the gradient of H dotted with the vector field. This must approach zero, and since the Liapunov function H is chosen so its gradient is never perpendicular to the vector field, the vector field itself must approach zero. In other words, the position vector must converge to the set of points at which the vector field is zero, which are the critical points for the flow. In the discrete-time stochastic version, namely the VRRW, it is the expectation of $H(\vec{v}(n))$ that increases, so by breaking down $H(\vec{v}(n))$ into a convergent martingale plus an increasing process, a similar argument goes through.

Definition 5.2 *For any vector $\vec{v}$, let $M_{\vec{v}}$ denote the Markov chain with transition matrix $M(\vec{v})$ given by*

$$M_{ij} = prob(j \to i) = \mathbf{R}_{ij}v_i/N_j. \tag{4}$$

Let $\pi(\vec{v})$ denote the stationary distribution for $M_{\vec{v}}$. (Later we will put a condition on $\vec{v}$ and $\mathbf{R}$ to ensure that $\pi(\vec{v})$ is unique.) It is easy to verify that $\pi(\vec{v})$ is given by

$$\pi(\vec{v})_i = v_i N_i / H(\vec{v}) \tag{5}$$

where

$$H(\vec{v}) = \sum v_i N_i$$

normalizes the expression so the coordinates sum to 1. Define $\pi(\vec{v})$ by (5) when it is not unique. Another way to write H is

$$H(\vec{v}) = \sum_{r,s} \mathbf{R}_{rs} v_r v_s. \tag{6}$$

In fact $H(\vec{v})$ is the potential function mentioned earlier; it always increases as $\vec{v}$ moves towards $\pi(\vec{v})$ as will now be demonstrated.

Definition 5.3 *Let* $face(\vec{v}) = \{\vec{w} \in \triangle : v_i = 0 \Rightarrow w_i = 0\}$ *be the unique face of* $\triangle$ *to which* $\vec{v}$ *is interior.*

Proposition 5.4 *For any* $\vec{v} \in \triangle$, $\vec{\nabla} H(\vec{v}) \cdot (\pi(\vec{v}) - \vec{v}) \geq 0$. *Furthermore, the following are equivalent:*

$$\begin{array}{ll} (i) & \vec{\nabla} H(\vec{v}) \cdot (\pi(\vec{v}) - \vec{v}) = 0 \\ (ii) & \vec{\nabla} H|_{face(\vec{v})} = \vec{0} \\ (iii) & \text{for those } i \text{ such that } v_i > 0\,,\ N_i \text{ are equal} \\ (iv) & \text{for all } i,\ v_i = \sum_j \mathbf{R}_{ij} v_i v_j / N_j \\ (v) & \pi(\vec{v}) = \vec{v} \end{array} \tag{7}$$

where $0/0 = 0$ *in (iv) by convention. We define a critical point to be one satisfying these equivalent conditions.*

The equivalent conditions in equations (7) are really just a form of the Chapman-Kolmogorov forward equation. For example, (iv) says that the expected fraction of time spent in state i in the future, gotten by summing the fraction of the time in state j times the probability of a transition from j to i over all j, is equal to the current value of v_i. This equation says nothing when $v_i = 0$. However, if we divide by v_i and replace the equality with an inequality, then theorem 5.11 says the inequality must hold even when $v_i = 0$ at any point $\vec{v}$ to which $\vec{v}(n)$ converges with non-zero probability. In terms of the forward equation, this inequality says that if v_i is to get to zero, the expected time at state i in the future must not exceed the current value of v_i. The assumption $H(\vec{v}) \neq 0$ really costs nothing since the actual process $\vec{v}(n)$ always satisfies $H(\vec{v}(n)) \geq 4I/(d^2+d)$ where $I = \min\{\mathbf{R}_{ij} : \mathbf{R}_{ij} > 0\}$. [Reason: there are $(d^2+d)/2$ edges so some edge is traversed at least $2/(d^2+d)$ of the time. If this edge connects vertices i and j then $\mathbf{R}_{ij} \geq I$ since it is nonzero and $v_i, v_j \geq 2/(d^2+d)$.]

Proof of proposition 5.4: Since $\vec{v} \in \Delta$, we know $H(\vec{v}) > 0$. For fixed i and j and constant c, consider the operation of increasing v_j by the quantity $cv_iv_j(N_j - N_i)$ and decreasing v_i by the same amount. If we let $c = 1/H(\vec{v})$ and do this operation simultaneously for every (unordered) pair i, j, then the resulting vector is $\pi(\vec{v})$: the next value of the i^{th} coordinate is given by

$$\begin{aligned} & v_i + (1/H(\vec{v}))(\textstyle\sum_j v_iv_jN_i - \sum_j v_iv_jN_j) \\ = \; & v_i + (1/H(\vec{v}))(v_iN_i - v_iH(\vec{v})) = \pi(\vec{v}). \end{aligned}$$

So an infinitesimal move towards $\pi(\vec{v})$ corresponds to doing these additions and subtractions simultaneously with an infinitesimal c. To show that this increases H, it suffices to show that for each unordered pair i, j, the value of H is increased, since H is smooth and therefore well approximated by its linearization near any point. So let i, j be arbitrary. Writing $\vec{v}^{(1)}$ for the new vector, we have

$$\begin{aligned} H(\vec{v}^{(1)}) &= \sum \mathbf{R}_{rs} {v_r}^{(1)} {v_s}^{(1)} \\ &= \sum_{r,s} \mathbf{R}_{rs} v_r v_s + 2 \sum_s \mathbf{R}_{is} c v_i v_j (N_i - N_j) v_s \end{aligned}$$

$$
\begin{aligned}
& +2\sum_r \mathbf{R}_{rj} c v_i v_j (N_j - N_i) \\
= \; & H(\vec{v}) + 2c v_i v_j (N_i - N_j)^2 \\
\geq \; & H(\vec{v})
\end{aligned}
$$

so H is nondecreasing. If there are any i and j for which $N_i \neq N_j$ and neither v_i nor v_j is zero, then H strictly increases.

Thus $(i) \Leftrightarrow (iii)$. Since

$$\vec{\nabla} H = (2N_1, \cdots, 2N_n), \tag{8}$$

and restricting to face($\vec{v}$) just throws out the coordinates i such that $v_i = 0$, it is easy to see that $(ii) \Leftrightarrow (iii)$. Assuming (iii), suppose the common value of the N_i is c. Then multiplying (iv) by c gives $\sum_j v_i v_j = c \cdot v_i$, so $(iii) \Rightarrow (iv)$. Now assume (iv). Letting $M_{\vec{v}}$ denote the matrix as well as the Markov chain, (iv) just says that $\vec{v}$ is stationary for $M_{\vec{v}}$. Then $\pi(\vec{v}) - \vec{v} = \vec{0}$ so (v) holds. And finally, $(v) \Rightarrow (i)$ trivially. □

Definition 5.5 *Let $\mathcal{C}$ be the set of critical points for H (i.e. those satisfying the equivalent conditions in proposition 5.4).*

Proposition 5.6 *The set $\mathcal{C}$ has finitely many connected components, each of which is closed and on each of which H is constant.*

Proof: By (7) (ii), $\mathcal{C}$ is the union over all $2^d - 1$ faces F of the sets $\mathcal{C}_F = \{\vec{v} : \vec{\nabla} H|_F(\vec{v}) = \vec{0}\}$. By (8) and the comment following, $\vec{\nabla} H|_F$ is linear on F, so $\mathcal{C}_F$ is a closed, convex, connected set. It is easy to see that H is constant on $\mathcal{C}_F$ by integrating $\vec{\nabla} H|_F$. The proposition follows since each connected component of $\mathcal{C}$ is the union of some of the $\mathcal{C}_F$. □

5.3 Convergence of $\vec{v}(n)$

We begin with a lemma stating that Liapunov function for the heuristic differential equation also serves as a non-decreasing potential for the expected value of the discrete-time process $\vec{v}(n)$, at least to within an error of order $1/n$.

Lemma 5.7 *Let $\mathcal{N}$ be a closed subset of the simplex, with $\mathcal{N} \cap \mathcal{C} = \emptyset$, and not containing any points where the Markov chain $M_{\vec{v}}$ is periodic or reducible (i.e. we require $M_{\vec{v}}$ to have a unique stationary distribution). Then there exist an N, L and $c > 0$ such that for any $n > N$, $\mathbf{E}(H(\vec{v}(n+L)) \mid \vec{v}(n)) > H(\vec{v}(n)) + c/n$ whenever $\vec{v}(n) \in \mathcal{N}$.*

[Note: The non-degeneracy condition on $M_{\vec{v}}$ is required to rule out cases like the following. Suppose $\mathbf{R}$ is the incidence matrix for a sparse graph and some $\vec{v} \in \mathcal{N}$ has enough coordinates equal to zero so that the non-zero coordinates represent a subset of the vertices of the graph that is not connected. Then $M_{\vec{v}}$ has at least two persistent sets of states between which no transitions can occur. So the stationary distribution is not unique and there is no guarantee that the behaviour over the next block of steps is correctly described by equation (5).]

Proof: First we claim that the lemma is true with the Markov process $M_{\vec{v}}$ substituted for $\vec{v}(n)$. Here $\vec{v}$ can be any point in Δ and the claim is that the lemma holds if the transition probabilities are frozen at tim n. Formally, define a process $\{\tilde{Y}_j\}$ for $j < n + L$ by $\tilde{Y}_j = Y_j$ for $j \leq n$ and letting $\tilde{Y}$ be Markov with transition matrix $M(\vec{v})$ after time n. Define the twiddled variables $\vec{\tilde{S}}$ and $\vec{\tilde{v}}$ by puting twiddles on equations (1) and (2). By proposition 5.4, $(\pi(\vec{v}) - \vec{v}) \cdot \vec{\nabla} H$ is nonzero on $\mathcal{N}$ so by compactness it is bounded below by some c_0 on $\mathcal{N}$. Choosing N/L large enough, se can get $\vec{v}(n+L)$ arbitrarily close to $\vec{v}(n)$ when $n > N$. Then by differentiability of H we will have $H(\vec{v} + (L/(N+L))(\pi(\vec{v}) - \vec{v})) > c_1$ for any $c_1 < c_0$. By the Markov property, $(\vec{\tilde{S}}(n+L) - \vec{S}(n))/L$ approaches a point-mass at $\pi(\vec{v})$ in distribution as L

increases. In fact, the rate of convergence of $M^k\vec{w}$ to $\pi(\vec{v})$ is exponential and controlled by the second-largest eigenvalue of M. Since M varies continuously with $\vec{v}$ and the non-degeneracy hypothesis says that $\mathcal{N}$ contains no points where the second-largest eigenvalue is 1, the second-largest eigenvalue is bounded away from 1. It follows that we can pick a large enough L uniformly in $\vec{v}$ so that $\mathbf{E}(H(M_{\vec{v}}(n+L)) \,|\, \vec{v}(n)) > c_2/n$ for any $c_2 < c_1$, and the claim is established.

Now we couple the Markov chain $M_{\vec{v}}(n)$ to $\vec{v}(n)$ in such a way as to move identically for as long as possible. Formally, we define $\{\tilde{Y}_j\}$ and $\{Y_j\}$ on a common measure space so that if $Y_j = \tilde{Y}_j$ for all $j < n+k$ then

$$prob(Y_{n+k} \neq \tilde{Y}_{n+k}) \,|\, Y_{n+k-1} = i) = \sum_j (1/2)|R_{ij}v_j/N_i - M_{ij}(\vec{v}(n))|.$$

Picking $c < c_2$ and N/L large enough so that

$$(L^2/N)(L/N)|\vec{\nabla} H| < (c_2 - c)/N, \tag{9}$$

the coordinates of $\vec{v}$ cannot change by more than L/N in L steps, so the probability of an uncoupling is bounded by L^2/N. Then $\mathbf{E}|H(\vec{v}(n+L)) - H(M_{\vec{v}}(n+L))| < (c_2 - c)/N$ by (9), and combining this with the earlier claim proves the lemma. □

Theorem 5.8 *Suppose* $\mathbf{R}_{ij} > 0$ *for all* $i \neq j$. *Then*

$$d(\vec{v}(n), \mathcal{C}) \to 0 \text{ almost surely as } n \to \infty$$

where $d(\vec{v}, \mathcal{C})$ *as usual denotes* $\min\{|\vec{v} - \vec{w}| : \vec{w} \in \mathcal{C}\}$. *In particular, if* $\mathcal{C}$ *is discrete then* $\vec{v}(n)$ *converges almost surely.*

Idea of proof: On any set $\mathcal{N}$ away from $\mathcal{C}$, lemma 5.7 says the expected value of $H(\vec{v}(n))$ grows, provided you sample at time intervals of size L. Actually, $H(\vec{v}(n))$ itself increases for n sufficiently large because the differences between $H(\vec{v}(n+L))$ and

$\mathrm{E}(H(\vec{v}(n+L))\,|\,\vec{v}(n))$ form a convergent martingale. The rate of increase is of order $1/n$ but so is the rate of change in position of $\vec{v}(n)$. So if $\vec{v}(n)$ goes from one given point of $\mathcal{N}$ to another, $H(\vec{v}(n))$ increases by an amount independent of time. The only way it can decrease again is for $\vec{v}(n)$ to leave $\mathcal{N}$ at a place where H is large and re-enter where H is small. The effect of such a possibility can be made arbitrarily small because H is constant on the connected components of $\Delta \setminus \mathcal{N}$.

Proof: Since the connected components, $\mathcal{C}_i, \ldots \mathcal{C}_k$ of $\mathcal{C}$ are closed, we have $m = \min\{d(\mathcal{C}_i, \mathcal{C}_j)\} > 0$. Pick any $r < m/3$. Let

$$\mathcal{N}_1{}^i = \{\vec{v} : d(\vec{v}, \mathcal{C}_i) < r\} \tag{10}$$

$$\mathcal{N}_1 = \Delta \setminus \bigcup_{i=1}^{k} \mathcal{N}_1{}^i. \tag{11}$$

Note that

$$i \neq j \Rightarrow d(\mathcal{N}_1{}^i, \mathcal{N}_1{}^j) > r. \tag{12}$$

By the preceding lemma with $\mathcal{N} = \mathcal{N}_1$, we can find c_1, L_1, N_1 for which $n \geq N_1$ implies $\mathrm{E}(H(\vec{v}(n+L))\,|\,\vec{v}(n)) \geq H(\vec{v}(n)) + c/n$. Pick any $L' > L_1$ and define

$$\mathcal{N}_2{}^i = \mathcal{N}_1{}^i \cap \{\vec{v} : |H(\vec{v}) - H(\mathcal{C}_i)| < rc/2L'\} \tag{13}$$

$$\mathcal{N}_2 = \Delta \setminus \bigcup_{i=1}^{k} \mathcal{N}_2{}^i. \tag{14}$$

Figure 2 gives an example of these definitions when $d = 3$; the heavy lines are the boundary of $\mathcal{N}_1$ and the lighter lines are the boundary of $\mathcal{N}_2$.

Apply the lemma to $\mathcal{N}_2$ to get N_2, c_2 and L_2. Define the process $\{\vec{u}(n)\}$ that samples $\vec{v}(n)$ at intervals of L_1 on $\mathcal{N}_1$ and L_2 elsewhere, by

$$\vec{u}(n) = \vec{v}(f(n))$$

where

$$\begin{aligned} f(1) &= \max\{N_1, N_2\} \text{ and} \\ f(n+1) &= \begin{cases} f(n) + L_1 \text{ if } \vec{v}(f(n)) \in \mathcal{N}_1 \\ f(n) + L_2 \text{ if } \vec{v}(f(n)) \in \mathcal{N}_2 \setminus \mathcal{N}_1 \end{cases} . \end{aligned}$$

Clearly, $\vec{u}(n)$ converges if and only if $\vec{v}(n)$ converges. Letting $U(n) = H(\vec{u}(n))$, we can write $U(n) = M(n) + A(n)$ where $\{M(n)\}$ is a martingale and $\{A(n)\}$ is a predictable process (take the filtration as $\mathcal{F}_n = \sigma\{Y_j : j \leq f(n)\}$). The key properties needed are

$$M(n) \text{ converges almost surely} \tag{15}$$

$$A(n+1) \geq A(n) + c/n \text{ if } \vec{u}(n) \in \mathcal{N}_1 \tag{16}$$

$$A(n+1) \geq A(n) \text{ if } \vec{u}(n) \notin \mathcal{N}_1. \tag{17}$$

To verify (15), note that $|U(n+1) - U(n)| \leq \sup|\vec{\nabla} H| \max\{L_1, L_2\}/f(n) = O(1/n)$. Then $|M(n+1) - M(n)| = O(1/n)$ as well, so $M(n)$ converges in L^2, hence almost surely. Properties (16) and (17) are evident from the construction.

The next thing to show is that with probability 1, $\vec{u}(n) \in \mathcal{N}_2{}^a$ infinitely often for at most one a. Consider any sample path $\vec{u}(1), \vec{u}(2), \ldots$. For $n < t$, define the event $\mathcal{B}(a, b, n, t, \omega)$ to occur if

$$\vec{u}(n) \in \mathcal{N}_2{}^a \text{ and } \vec{u}(t) \in \mathcal{N}_2{}^b \text{ with } \vec{u}(i) \in \mathcal{N}_2 \text{ for } n < i < t. \tag{18}$$

If $\mathcal{B}(a, b, n, t, \omega)$ occurs, let

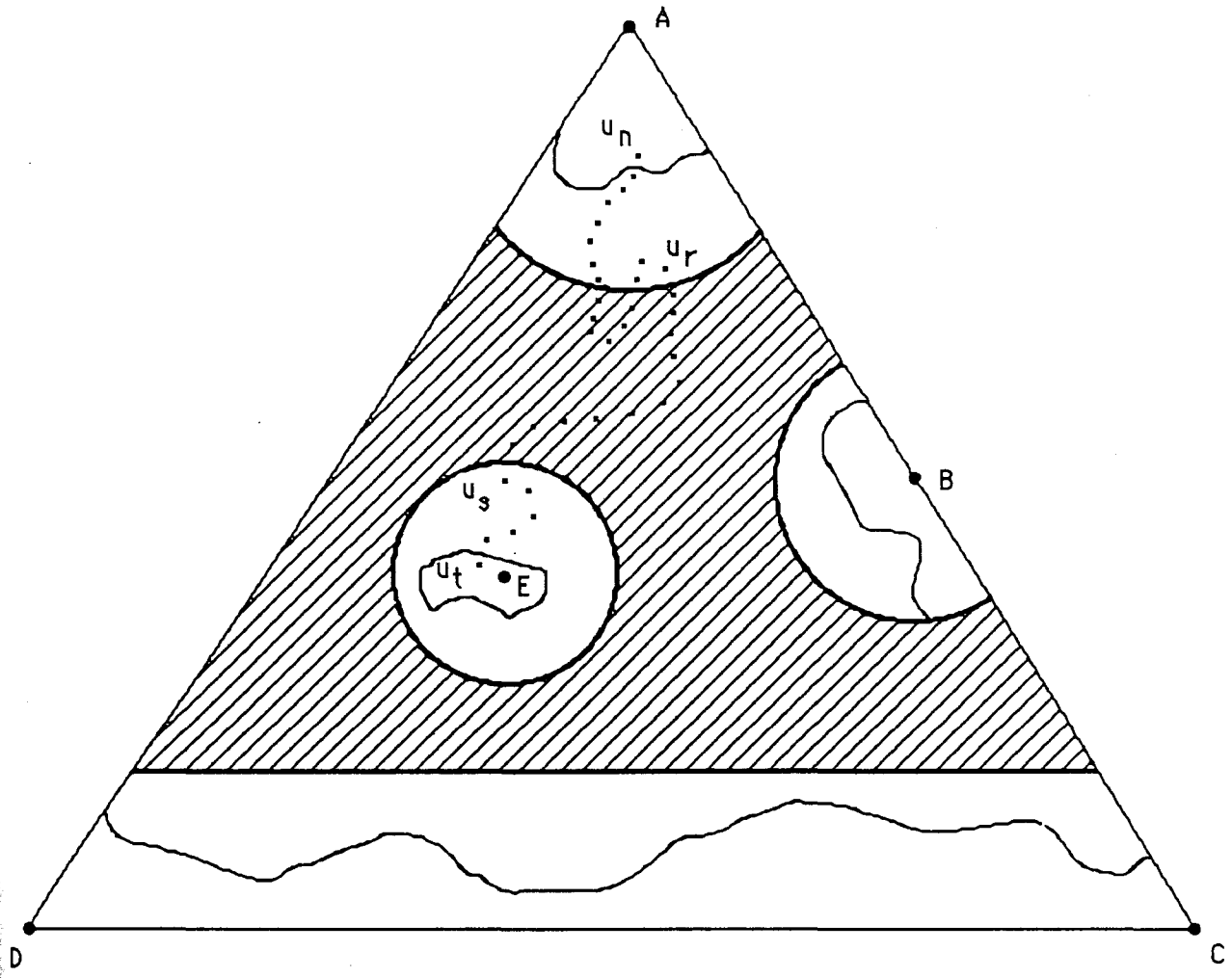
$$r = \max\{i : n \leq i < t \text{ and } \vec{u}(i) \in \mathcal{N}_1{}^a\} \text{ and}$$

$$s = \max\{i : n \leq i < t \text{ and } \vec{u}(i) \notin \mathcal{N}_1{}^b\} + 1$$

The dotted path in figure 6 gives an example of this. By (16) and (17) we have

$$A(i+1) - A(i) \geq c/i \text{ for } r < i < s$$

$$A(i+1) - A(i) \geq 0 \text{ for } n < i < t.$$



$\mathcal{C} = \{A\} \cup \{B\} \cup \overline{CD} \cup \{E\}$

$\mathcal{N}_1$ is the shaded region

FIGURE 6

Then

$$\begin{aligned}
& A(t) - A(n) \\
= \; & [A(t) - A(s)] + [A(s) - A(r+1)] \\
& + [A(r+1) - A(n+1)] + [A(n+1 - A(n)] \\
\geq \; & 0 + \left(\sum_{i=r+1}^{s-1} c/i \right) + 0 - L_2/n \\
= \; & O(1/n) + (c/L_1) \sum_{i=r}^{s-1} L_1/i \\
\geq \; & O(1/n) + (c/L_1) \sum_{i=r}^{s-1} |\vec{u}(i+1) - \vec{u}(i)| \\
\geq \; & O(1/n) + (c/L_1) |\vec{u}(s) - \vec{u}(r)| \\
> \; & O(1/n) + rc/L_1
\end{aligned}$$

by (12). Now $U(t) - U(n) \leq H(\mathcal{C}_b) - H(\mathcal{C}_a) + rc/L'$ by the construction of $\mathcal{N}_2$. So $M(t) - M(n) \leq H(\mathcal{C}_b) - H(\mathcal{C}_a) + rc/L' - rc/L_1 + O(1/n)$. If $H(\mathcal{C}_b) \leq H(\mathcal{C}_a)$, the choice of r guarantees that this expression is strictly negative and bounded away from 0 for large n. Therefore if $M(n)(\omega)$ converges, then $\mathcal{B}(a, b, n, t, \omega)$ happens only finitely often for a, b such that $H(\mathcal{C}_b) \leq H(\mathcal{C}_a)$. But then it happens only finitely often for any $a \neq b$, since $\vec{u}$ can make only $k-1$ successive transitions from $\mathcal{N}_2{}^a$ to $\mathcal{N}_2{}^b$ with $H(\mathcal{C}_b) > H(\mathcal{C}_a)$. Thus the almost sure convergence of $M(n)$ implies that $\vec{u}(n) \in \mathcal{N}_2{}^a$ infinitely often for at most one a, almost surely.

In other words, transitions between small neighborhoods of $\mathcal{C}_i$ and $\mathcal{C}_j$ eventually cease for $i \neq j$. It remains to show that $\vec{v}(n)$ may not remain far from $\mathcal{C}$, nor may it keep oscillating between a small neighborhood of $\mathcal{C}_i$ and a set bounded away from $\mathcal{C}$. To do this, require now that $r < m/6$. With $\mathcal{N}_1$ and $\mathcal{N}_2$ defined as before, define $\mathcal{N}_3$ by (11) with $2r$ in place of r. Since $2r < m/3$, equation (12) holds with $\mathcal{N}_3$ in place of $\mathcal{N}_1$. An argument identical to the one above now shows that with probability 1 there

are only finitely many values of n, i and t with $n < i < t$ for which

$$\vec{u}(n)(\omega) \in \mathcal{N}_2{}^a, \vec{u}(i)(\omega) \in \mathcal{N}_3 \text{ and } \vec{u}(t)(\omega) \in \mathcal{N}_2{}^a.$$

[The argument again: $A(i)$ is nondecreasing when $\vec{u}(n) \in \mathcal{N}_1{}^a$ and increases by at least the fixed amount rc/L_1 each time $\vec{u}$ makes the transit from $\mathcal{N}_1{}^a$ to $\mathcal{N}_3$. The increase in A is greater than the greatest difference in values of H taken at two points of $\mathcal{N}_2{}^a$, so the martingale M must change by at least $rc/L_1 - rc/L'$ during every transit. Since M converges, this happens finitely often.]

Similarly, we find that the event $\{\omega : \vec{u}(t,\omega) \in \mathcal{N}_1 \text{ for all } t > n\}$ has probability 0 for each n. Putting all of this together, we find that for any small r there is precisely one a for which $\vec{u}(n) \in \mathcal{N}_1{}^a$ infinitely often. But then for any r, $\mathcal{N}_3$ stops being visited, so letting $r \to 0$ proves the theorem. □

Corollary 5.9 *Let* $\mathbf{R}$ *be any symmetric matrix. Then there are symmetric matrices arbitrarily close to* $\mathbf{R}$ *(in the sup norm on entries) for which* $\vec{v}(n)$ *converges.*

Proof: For Arbitrary symmetric $\mathbf{R}$, add a small number to each entry symmetrically so it is non-zero. Then add a small multiple of the diagonal matrix so that it and all its minors are invertible and hence there is only one critical point per face. Then apply theorem 5.8. □

To point out what can go wrong when $\mathbf{R}$ has zeros off the diagonal, suppose $\mathbf{R}_{12} = \mathbf{R}_{21} = 0$. Then $H(x, 1-x, 0, \ldots, 0) = \mathbf{R}_{11}x^2 + \mathbf{R}_{22}(1-x)^2$. Now if $\vec{v}(n)$ is very near $(x, 1-x, 0, \ldots, 0)$ for some $x \in (0,1)$ then the VRRW is likely to stay at vertex 1 for a while and then leak over to vertex 2 by way of some intermediate vertex j with $\mathbf{R}_{1j} \neq 0 \neq \mathbf{R}_{j2}$. If $\mathbf{R}_{12}$ were greater than zero, the leakage could occur directly and the leakage back and forth would occur often enough to be quantitatively taken into account. In particular, the time spent at vertices 1 and 2 would be given by the

stationary distribution $\pi(x, 1-x, 0, \ldots, 0)$. But since $\mathbf{R}_{12} = 0$, the Markov chain $M_{\vec{v}}$ for $\vec{v} = (x, 1-x, 0, \ldots, 0)$ has transition matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ on the first two vertices, so any distribution is stationary. In particular, the value of $H(\vec{v}(n))$ may decrease if x is near 1 and leakage to vertex 2 occurs. Then the argument for theorem 5.8 breaks down entirely, since $\vec{v}(n)$ may continue to wander around $\triangle$ indefinitely, getting near $(1-\epsilon, \epsilon, 0, \ldots, 0)$ as H gets large and then "recharging" H to a smaller value by leaking to vertex 2 for a while. While this behaviour can sometimes be ruled out on an ad hoc basis, a general theorem is lacking.

5.4 Local behaviour of the Liapunov function at the point of convergence

Suppose $\vec{p}$ is a point for which $prob(\vec{v}(n) \to \vec{p}) > 0$. It seems likely that H must have a local maximum at $\vec{p}$, or a weak local maximum if $\vec{p}$ is not an isolated point of C. In this section and the next two results are established in this direction. It will be shown that if $\vec{p}$ is in the interior of $\triangle$ and $\mathbf{R}$ satisfies a non-degeneracy condition, then H must have a local maximum at $\vec{p}$ (necessarily a strict one, since H is quadratic and its matrix representation is non-degenerate). If $\vec{p}$ is on a proper face of $\triangle$, the linear approximation to H at $\vec{p}$ is shown to have a weak maximum at $\vec{p}$. To illustrate when these results apply, consider the case $d = 2$ where H can be thought of as a function from $[0, 1]$ to itself. Then the critical points are $0, 1$, and anywhere where H' vanishes. The above conditions would then say that $H''(p) \leq 0$ for $p \in (0, 1)$, $H'(0) \geq 0$ and $H'(1) \leq 0$ were necessary for convergence to p, 0 or 1 respectively. In higher dimensions, these results do not address the case where H fails to be a maximum on $face(\vec{v})$ but is a maximum in every direction normal to $face(\vec{v})$.

To state these conditions formally, let $e_1, \ldots, e_d$ be the standard basis vectors.

Definition 5.10 *A critical point $\vec{p}$ on a proper face of Δ is a* linear non-maximum *if*

$$\vec{\nabla} H \cdot (e_k - e_j) > 0 \text{ for some } e_k \notin face(\vec{p}), e_j \in face(\vec{p}). \tag{19}$$

Let $W = \{\vec{w} \in \mathbf{R}^d : \sum_i w_i = 0\}$. Then we have

Theorem 5.11 *Suppose $\vec{p}$ is on a proper face of Δ and is a linear non-maximum. Then $prob(\vec{v}(n) \to \vec{p}) = 0$. In fact there is some neighborhood of $\vec{p}$ that $\vec{v}(n)$ exits almost surely (i.e. even when $\vec{v}(n)$ has a limit with a non-atomic distribution, $\vec{p}$ is not in the support. If $\vec{v}(n)$ is known to converge, we may say that $\vec{v}(n)$ visits this neighborhood finitely often rather than just exiting it almost surely.)*

Remark: The condition on the eigenvalues of $diag(\vec{p})\mathbf{R}$ is only needed for a technical lemma, namely the Sternberg linearization theorem (5.24 below). I do not believe it is really necessary and hope to remove it in the near future.

Theorem 5.12 *Suppose that $\mathbf{R}$ is nonsingular and let $\vec{p}$ be the critical point in the interior of Δ. Suppose in addition that the eigenvalues of $diag(\vec{p})\mathbf{R}$ are distinct and linearly independent over the rationals, where $diag(\vec{p})$ is the diagonal matrix with i,i entry equal to p_i. Then $prob(\vec{v}(n) \to \vec{p}) = 0$ whenever $\vec{p}$ fails to be a maximum for H. This happens if and only if the matrix $\mathbf{R}$ has only one positive eigenvalue.*

Corollary 5.13 *Let $\mathbf{R}$ be any symmetric non-singular matrix and let $\vec{p}$ be a critical point for the vertex-reinforced random walk defined from $\mathbf{R}$. If $\vec{p}$ is in the interior of Δ and is not a maximum for H, then there is a neighborhood $\mathcal{N}$ of $\vec{p}$ and there are symmetric matrices arbitrarily close to $\mathbf{R}$ (in the sup norm on entries) for which the process $\vec{v}(n)$ cannot converge to any point in $\mathcal{N}$.*

Proof of corollary 5.13: Since $\vec{p} = \lambda \mathbf{R}^{-1}(1, \ldots, 1)^T$ by proposition 5.4 (iii) and $\mathbf{R}$ is non-singular, $\vec{p}$ is an isolated critical point. By the same reasoning, small perturbations of $\mathbf{R}$ have precisely one critical point in a neighborhood of $\vec{p}$. A generic perturbation of $\mathbf{R}$ will satisfy the hypotheses of theorem 5.12. □

Before proving theorem 5.11, we need to translate the condition (19) into a more usable form. If $\vec{p}$ is a critical point then equation (7) (iii) says that the N_i for i such that $p_i > 0$ have a common value, λ. Assuming (19) for a given e_k and using equation (8) for $\vec{\nabla} H$, we see that $N_k > N_j = \lambda$. So

$$\sum_j \mathbf{R}_{kj} p_j / N_j = \sum_{p_j > 0} \mathbf{R}_{kj} p_j / \lambda = N_k / \lambda = 1 + b \tag{20}$$

for some $b > 0$, k such that $p_k = 0$. A skeleton for the proof of theorem 5.11 will be as follows.

Let $\vec{p}$ be as above. For any $\delta > 0$, the occupation vector $(\vec{S}(n + \delta n) - \vec{S}(n))/\delta n$ can be forced to be in a neighborhood of $\vec{p}$ by assuming that $\vec{v}(n)$ stays in a smaller neighborhood of $\vec{p}$ (proposition 5.14 below). This will force $\sum_i (S_i(n + \delta n) - S_i(n)) \mathbf{R}_{ik} v_k / (1 + \delta) N_k(n)$ to be greater than $\delta S_k(n)$ (proposition 5.15 below), where this quantity represents in some sense the expected number of visits to vertex k between times n and $n + \delta n$. This implies that $\ln(v_k(n))$ grows at a rate that is bounded below when $\vec{v}(n)$ is in a small enough neighborhood of $\vec{p}$, so $\vec{v}(n)$ cannot stay in this neighborhood for ever.

To avoid bogging down in trivialities assume that $\vec{S}(t)$ is defined for non-integral t by linear interpolation with $\vec{v}(t) = \vec{S}(t)/t$. The approximations will be verified only in the cases where the arguments are integral; it is always possible to choose epsilons and deltas a little bit smaller to compensate for the roundoff errors.

Theorem 5.11 is proved as a sequence of propositions. The first of these says that if $\vec{v}(n)$ always stays in some neighborhood $\mathcal{N}$ then the fractional occupation vector for

the walk between the times n and $n + \delta n$

$$(\vec{S}(n + \delta n) - \vec{S}(n))/\delta n$$

must stay in a slightly larger neighborhood, $\mathcal{N}_1$. Actually, $\mathcal{N}_1$ and δ are given, and $\mathcal{N}$ is found so that the above is true uniformly in n.

Proposition 5.14 *Fix $\vec{p}$ and let $\mathcal{N}_1$ be a neighborhood of $\vec{p}$. For any $\delta > 0$ there is a neighborhood $\mathcal{N}$ of $\vec{p}$ included in $\mathcal{N}_1$ such that for all $n > 1/\delta$, the two conditions*

$$(i) \quad \vec{v}(n) \in \mathcal{N} \qquad \textit{and}$$
$$(ii) \quad \vec{v}(n + \delta n) \in \mathcal{N}$$

imply

$$(iii) \quad (\vec{S}(n + \delta n) - \vec{S}(n))/\delta n \in \mathcal{N}_1 \quad .$$

Proof: The term in (ii) is a convex combination of the terms in (i) and (iii):

$$\vec{v}(n + \delta n) = \frac{1}{1+\delta}\vec{v}(n) + \frac{\delta}{1+\delta}\frac{\vec{S}(n + \delta n) - \vec{S}(n)}{\delta n}.$$

Solving for the last term gives

$$\frac{\vec{S}(n + \delta n) - \vec{S}(n)}{\delta n} = \vec{v}(n + \delta n) + (1/\delta)(\vec{v}(n + \delta n) - \vec{v}(n)).$$

Since $\mathcal{N}_1$ contains some ball centered at $\vec{p}$, we can choose $\mathcal{N}$ to be a ball whose radius is $(1 + 2/\delta)^{-1}$ times as big. □

The next step is to show that if the fractional occupation vector between times n and $n + \delta n$ is in an appropriate neighborhood $\mathcal{N}_1$ of $\vec{p}$, then the number of expected transitions to vertex k is large. To elaborate, k is the vertex referred to in the definition of a linear non-maximum 5.10. The "expected number of transitions" means the sum of

$prob(\vec{v}(i+1) = k|\vec{v}(i))$ over all i between n and $n+\delta n$. This is a random variable, but it is shown to be always sufficiently large under the hypotheses of the proposition (no set of measure needs to be excluded). The quantity on the right hand side of equation 21 is sufficiently large because it makes proposition 5.17 true.

Proposition 5.15 *Let $\vec{p}$, k be such that (20) holds and let $\vec{D}$ be any vector function of n. Then there is an $\epsilon > 0$ and a neighborhood $\mathcal{N}_1 = \{\vec{v} \in \Delta : |\vec{v} - \vec{p}| < \epsilon\}$ such that for all $\delta > 0$ and for all n, the conditions $\vec{v}(n) \in \mathcal{N}_1$ and $(D_i(n+\delta n) - D_i(n))/\delta n \geq p_i - \epsilon$ for all i imply*

$$\sum_i (D_i(n+\delta n) - D_i(n))\mathbf{R}_{ik} v_k(n)/(1+\delta)N_i(n) > \frac{1+b/2}{1+\delta} S_k(n). \qquad (21)$$

Proof: As $\epsilon \to 0$, $1/n$ times the left-hand side converges to $\delta p_k(n)/(1+\delta)\sum_i p_i(n)\mathbf{R}_{ik}/N_i(n)$ $= \delta p_k(n)(1+b)/(1+\delta)$ while $1/n$ times the right-hand side converges to $\delta p_k(n)(1+b/2)/(1+\delta)$. Since the convergence is uniform in δ, the result follows. □

To make use of this proposition, we need to show that the number of visits to vertex k will be large when its expected value is large. To do this the following facts about sums of Bernoulli variables are required.

Proposition 5.16 *Let $b > 0$ and $\epsilon > 0$ be given. let $\{B_\alpha\}$ be a collection of independent Bernoulli random variables with $\mathbf{E}(\sum_\alpha B_\alpha) \geq (1+b)L$.*

(1) There exists an L_0 such that whenever $L > L_0$, we have $prob(\sum_\alpha B_\alpha/L > 1 + b/2) > 1 - \epsilon$.

(2) For any $c < 1 + b$ there is an $\epsilon > 0$ such that whenever $L > 1$ and $\mathcal{A}$ is an event with $prob(\mathcal{A}) > 1 + \epsilon$ then $\mathbf{E}(\sum_\alpha B_\alpha \,|\, \mathcal{A}) > cL$.

Proof: Let $p_\alpha = prob(B_\alpha = 1)$, so the hypothesis says that $\sum_\alpha p_\alpha \geq (1+b)L$; assume without loss of generality that the sum is in fact equal to $(1+b)L$. The variance of B_α is given by $p_\alpha - {p_\alpha}^2$, so the variance of $\sum B_\alpha$ is finite and less than $(1+b)L$. Then the variance of $\sum B_\alpha/L$ is less than $(1+b)/L$, so by Tschebysheff's inequality, $prob(\sum B_\alpha < 1 + b/2) < 4(1+b)/Lb^2$. So we can choose $L = 4(1+b)/b^2\epsilon$.

For the second part, we have

$$\mathbf{E}(\sum B_\alpha/L \,|\, \mathcal{A}) = \mathbf{E}(\sum B_\alpha/L) - \mathbf{E}(\sum B_\alpha \mathbf{1}_{\overline{\mathcal{A}}}/L)/prob(\mathcal{A}).$$

But letting F be the distribution function for $\sum B_\alpha/L$, we have

$$\begin{aligned} \mathbf{E}(\sum B_\alpha \mathbf{1}_{\overline{\mathcal{A}}}/L) &\leq \int_{F^{-1}(1-\epsilon)}^{\infty} prob(\sum B_\alpha/L > y)dy \\ &\leq \int_{1+b+\left(\frac{1+b}{\epsilon L}\right)^{1/2}}^{\infty} \frac{1+b}{Ly^2}\,dy \end{aligned}$$

by Tschebysheff's inequality, using the bound on the variance calculated in the preceding paragraph. Since this is at most $\sqrt{\epsilon/(1+b)}$, the result follows. □

The next propostion uses the above calculation to show that the geometric mean of the fractional occupation $v_k(n)$ at vertex k increases at a fixed rate.

Proposition 5.17 *Let $b > 0$ and $\gamma \in (0, 1/2)$. There are an $L_0(b)$ and an $F(b,\gamma)$ such that the following holds: if $\{B_\alpha\}$ are a collection of independent Bernoulli random variables with $\mathbf{E}(\sum_\alpha B_\alpha) \geq (1+b)L$ and $L > L_0$, then $\mathbf{E}(\ln(\gamma \sum_\alpha B_\alpha/L + 1 - \gamma)) > F(b,\gamma) > 0$.*

Proof: Pick $k > 0$ so that $\ln(1+x) > kx$ on $(1, 1+b/2)$. Let $\epsilon < \min(\frac{1}{10}, \frac{9kb/20}{2\ln(1/2)})$. Apply proposition 5.16 to find an L_0 for these values of b and ϵ. Then for any $L > L_0$ the random variable $\sum_\alpha B_\alpha/L$ stochastically dominates the variable θ which is 0 with

probability ϵ and $1 + b/2$ with probability $1 - \epsilon$. Then

$$\begin{aligned} & \mathbf{E}(\ln(\gamma \textstyle\sum_\alpha B_\alpha/L + 1 - \gamma)) \\ \geq\ & \mathbf{E}(\ln(\gamma\theta + 1 - \gamma)) \\ =\ & \epsilon \ln(1-\gamma) + (1-\epsilon)\ln(1 + \gamma b/2) \qquad (22) \\ \geq\ & \epsilon(2\gamma \ln(1/2)) + (9/10)(k\gamma b/2) \\ \geq\ & \gamma(9kb/20 - 2\epsilon \ln(1/2)) \end{aligned}$$

so letting $F(b,\gamma) = \gamma(9kb/20 - 2\epsilon\ln(1/2))$, the proposition is proved. □

Proof of theorem 5.11: By hypothesis, condition (19) and hence (20) hold. Pick ϵ and $\mathcal{N}_1$ according to proposition 5.15. Pick any $\delta < b/4$ where b is as in (20). Now apply proposition 5.14 to get an $\mathcal{N} \subseteq \mathcal{N}_1$ with the appropriate properties. Fix n and define a set of Bernoulli random variables $\{B_{i,r}\}$ as follows.

> Let $\tau_{i,r} \leq \infty$ be the r^{th} time after n that $Y_j = i$, so formally $\tau_{i,0} = n$ and $\tau_{i,r+1} = \inf\{j > \tau_{i,r} : Y_j = i\}$. Let $B_{i,r}$ be independent and Bernoulli with
>
> $$prob(B_{i,r} = 1) = \mathbf{R}_{ki} v_k(n)/(1+\delta)N_i(n) \qquad (23)$$
>
> and coupled to the variables $\{Y_i\}$ so that if $B_{i,r} = 1$ and $\tau_{i,r} \leq n + \delta n$ then $Y_{\tau_{i,r}+1} = k$.

To verify that this construction is possible, check that the probability of a transition from vertex i to vertex k never drops below the quantity in (23):

$$\begin{aligned} prob(Y_{\tau_{i,r}+1} = k \,|\, \mathcal{F}_{\tau_{i,r}}) &\geq (n/\tau_{i,r})\mathbf{R}_{ki} v_k(n)/N_i(n) \\ &\geq (1/(1+\delta))\mathbf{R}_{ki} v_k(n)/N_i(n) \end{aligned}$$

for $\tau_{i,r} < n + \delta n$.

Now consider the subcollection $\{B_{i,r} : r \leq \delta n(p_i - \epsilon)\}$. Letting A denote the index set for this subcollection and writing α for a generic pair i, r in A, we have

$$\mathbf{E}(\sum_A B_\alpha) = \sum_i \delta n(p_i - \epsilon)\mathbf{R}_{ki} v_k(n)/(1+\delta)N_i(n).$$

(As promised, we are ignoring the roundoff error when $\delta n(p_i - \epsilon)$ is not an integer.) By the construction of ϵ from proposition 5.15, this quantity is at least $(1+b/2)S_k(n)/(1+\delta)$. By the second part of proposition 5.16, r can be chosen independently of $S_k(n)$ so that if $\mathcal{A}$ is any event with probability at least $1-r$ then

$$\mathbf{E}(\sum_A B_\alpha \mid \mathcal{A}) \geq (1 + b/4)S_k(n)/(1+\delta). \tag{24}$$

We now show that $S_k(n) \to \infty$ if $\vec{v}(n)$ does not leave $\mathcal{N}$. Assume to the contrary. Then there is an event $\mathcal{A} \in \mathcal{F}_N$ for some N such that $prob(\mathcal{A}) > 1 - r$ and $\mathcal{A}$ is the event

$$\vec{v}(n) \in \mathcal{N} \text{ for all } n \geq N \text{ and } S_k(n) = S_k(N) \text{ for all } n \geq N.$$

By proposition 5.14, $S_i(n+\delta n) - S_i(n) \geq (p_i - \epsilon)\delta n$ for all $n \geq N$ when $\mathcal{A}$ occurs, so by the coupling of $\{Y_i\}$ and $\{B_{i,r}\}$, $S_k(n+\delta n) \neq S_k(n)$ whenever $\sum_A B_\alpha > 0$. Then equation (24) contradicts the fact that $\sum_A B_\alpha = 0$ on $\mathcal{A}$.

Now to show that $\vec{v}(n)$ eventually leaves $\mathcal{N}$, assume to the contrary. For an S_0 to be determined later, we pick N so that the event that $\vec{v}(n) \in \mathcal{N}$ for all $n \geq N$ and $S_k(N) > S_0$ has probability at least $1-r$. Call this event $\mathcal{A}$. For $n \geq N$, calculate

$$\begin{aligned}
& \mathbf{E}(\ln(v_k(n+\delta n)) \mid \mathcal{A}, \mathcal{F}_n) \\
= \; & \mathbf{E}(\ln(v_k(n) + (\delta/(1+\delta))[(S_k(n+\delta n) - S_k(n))/(\delta n) - v_k(n)]) \mid \mathcal{A}, \mathcal{F}_n) \\
\geq \; & \mathbf{E}(\ln(v_k(n)/(1+\delta) + (\delta/(1+\delta))\textstyle\sum_A B_\alpha/\delta n) \mid \mathcal{F}_n) \\
= \; & \ln(v_k(n)) + \mathbf{E}(\ln(1/(1+\delta)) + (\delta/(1+\delta))\textstyle\sum_A B_\alpha/\delta S_k(n)) \mid \mathcal{F}_n).
\end{aligned} \tag{25}$$

Using proposition 5.17 with $\gamma = \delta/(1+\delta)$, this is bounded below by $\ln(v_k(n)) + F(b, \delta/(1+\delta)) > 0$ as long as $\delta S_k(n)$ is larger than some L_0 depending only on b. But since $S_k(n)$ grows without bound when $\vec{v}(n)$ stays in $\mathcal{N}$, N can be chosen large enough so that S_0 can be taken to be at least L_0/δ and the hypotheses of proposition 5.17 will be satisfied. This gives a contradiction, since $\ln(v_k(n))$ is bounded above by zero, but its expected value increases without bound on a set of non-zero measure, when sampled at consecutive times, $N, (1+\delta)N, (1+\delta)^2N, \ldots$. □

5.5 Proof of theorem 5.12

The last equivalence in theorem 5.12 is easy. The matrix $\mathbf{R}$ can be viewed as a symmetric bilinear form whose quadratic form gives H when restricted to W. For $\vec{w} \in W$,

$$\mathbf{R}(\vec{w}, \vec{p}) = \vec{w}^T \mathbf{R} \vec{p} = \vec{w} \cdot \lambda \cdot (1, \ldots, 1) = 0$$

where λ is the common value of the N_i. Then

$$\mathbf{R}(\vec{w} + c\vec{p}, \vec{w} + c\vec{p}) = \mathbf{R}(\vec{w}, \vec{w}) + \mathbf{R}(c\vec{p}, c\vec{p}) = H(\vec{w}) + c^2\lambda$$

so the quadratic form $\mathbf{R}(\vec{v}, \vec{v})$ decomposes into the sum of H and a positive form on the one-dimensional subspace spanned by $\vec{p}$. Then $\mathbf{R}$ has precisely one more positive eigenvalue then the quadratic form H. Since $\mathbf{R}$ has no zero eigenvalues, this means that H will have a strict maximum when it has a maximum, which will be when it has no positive eigenvalues. □

To begin proving the main part of the theorem, we start with a lemma that allows us to shift our focus from H to the vector field $\pi - I$ which maps $\vec{v}$ to $\pi(\vec{v}) - \vec{v}$.

Lemma 5.18 *Let $T : W \to W$ be the linear operator approximating $\pi(\vec{v}) - \vec{v}$ near $\vec{p}$, so*

$$\pi(\vec{p} + \vec{w}) - (\vec{p} + \vec{w}) = \vec{p} + T(\vec{w}) + O(|\vec{w}|^2). \tag{26}$$

Then T has real eigenvalues and T has a positive eigenvalue if and only if $\mathbf{R}$ has more than one positive eigenvalue.

Proof: First note that π is smooth on the interior of Δ, so T exists. We find a matrix M representing an extension of T to all of $\mathbf{R}^d$. Using formula (5) for π we have

$$M_{ij} = \frac{\partial}{\partial e_j}(\pi(\vec{v}))_i \Big|_{\vec{v}=\vec{p}} - \delta_{ij}$$

$$
\begin{aligned}
&= \frac{\partial}{\partial e_j}\frac{v_i N_i}{H(\vec{v})}\bigg|_{\vec{v}=\vec{p}} - \delta_{ij} \\
&= \left(\frac{\mathbf{R}_{i,j}v_j}{H} + \frac{\delta_{ij}N_i}{H} - \frac{v_i N_i \partial H/\partial e_j}{H^2}\right)\bigg|_{\vec{p}} - \delta_{ij} \\
&= \mathbf{R}_{ij}p_i/\lambda - 2p_i
\end{aligned}
$$

using the fact that all the N_i have a common value $\lambda = H(\vec{p})$.

Now the operator $\tilde{T}$ whose matrix is given by

$$\tilde{T}_{ij} = \mathbf{R}_{ij}p_i/\lambda$$

agrees with T on W since the difference is a matrix with constant rows, hence is zero on W. Letting $diag(\vec{p})$ be the diagonal matrix with i,i entry equal to p_i, we have $\tilde{T} = diag(\vec{p})\mathbf{R}/\lambda$. Since $\mathbf{R}$ is symmetric and $diag(\vec{p})$ is positive definite, we get that $\tilde{T}$ is diagonalizable with real eigenvalues and has the same signature as $\mathbf{R}$ (see [Or] theorem 6.23 and 6.24 p. 232). Since $\tilde{T}$ has $\vec{p}$ as a positive eigenvalue and W as an invariant subspace, it has one more positive eigenvalue than T and the conclusion follows. □

To finish proving theorem 5.12 we need to show that $\vec{v}(n)$ cannot converge to an interior point where the linear approximation T to the drift $\pi(\vec{v}) - \vec{v}$ has a positive eigenvalue. We follow the methods of chapter three. To turn the sequence $\{\vec{v}(n)\}$ into a randomized nonlinear urn scheme with tractable properties, we find a sequence of times $\tau_r, \tau_{r+1}, \ldots$ at which to sample, and consider instead the sequence $\{\vec{v}(\tau_n)\}$. Then in the terminology of chapter three, $\vec{G}(n)$ would be given by $\vec{v}(\tau_{n+1}) - \vec{v}(\tau_n)$. We then need to show that conditions (20) - (22) of chapter three are satisfied and to construct an η for which (23) of chapter three is satisfied. Then we can apply theorem 3.8.

This approach obscures the ideas more than is necessary, seeing as theorem 3.8 is just a way of verifying conditions (26) - (29) of chapter three. Since it is just as easy to verify them directly, we will do so. In addition, we would like to use $(\pi - I)\vec{v}(n)$

for $\overline{G}(n)$, but since this is only approximately true we must delve once more into the guts of the argument, which are therefore better exposed than hidden. We begin by restating what is to be shown. We must find a bounded function $\eta : \Delta \to \mathbf{R}$ and times $\tau_r, \tau_{r+1}, \ldots$ such that if $\eta(\vec{p}) = 0$ and if

$$S_n = \eta(\vec{v}(\tau_n)) \text{ and } X_n = S_n - S_{n-1} \tag{27}$$

then condition (32) below is satisfied. To state (32) we first state four other conditions.

$$\mathbf{E}({X_{n+1}}^2 + 2X_{n+1}S_n \,|\, \mathcal{F}_n) \geq b_1/n^2 \tag{28}$$

$$\mathbf{E}(X_{n+1}S_n \mathbf{1}_{|S_n|>c/n} \,|\, \mathcal{F}_n) \geq 0 \tag{29}$$

$$|X_n| \leq 1/n^\gamma \tag{30}$$

$$\mathbf{E}({X_{n+1}}^2 \,|\, \mathcal{F}_n) \leq b_2/n^2 \tag{31}$$

where

$$\begin{aligned} b_1, b_2, c &> 0 \\ \gamma &> 1/2 \\ \mathcal{F}_n &= \sigma\,(\text{all events up to time } \tau_n). \end{aligned}$$

Now the condition we want is that there is a neighborhood $\mathcal{N}$ of 0 such that for any ϵ there is an n sufficiently large so that

$$prob(\mathcal{B} \,|\, \mathcal{F}_n) > 1 - \epsilon \tag{32}$$

where $\mathcal{B} \subseteq \bigcap_{k\geq 0} \mathcal{B}_k$ and $\mathcal{B}_k \in \mathcal{F}_k$ is the event that either (28) - (31) are satisfied, or $S_n \notin \mathcal{N}$. Assuming (32) we sketch the argument that S_n cannot converge to 0. (For a more detailed version, read through the proof of theorem 3.5.)

First assume (28) - (31).

(A) Given any S_n, the probability of finding $|S_M| > k/\sqrt{n}$ for $k < \sqrt{b_1/2}$ and some $M > n$ is at least $1/2$: The expected square of S_{n+i}, stopped at the exit time τ of the interval $|x| < k/\sqrt{n}$, grows by at least $(b_1/n^2)prob(\tau > i)$ at step i, since the left-hand side of (28) is just $\mathbf{E}({S_{n+1}}^2 - {S_n}^2 \,|\, \mathcal{F}_n)$. If $prob(\tau = \infty) > 1/2$ then by summing b_1/n^2 we see that the expected square of the stopped process eventually exceeds ${S_n}^2 + b_1/2n$. But by (30), the stopped process never leaves an interval slightly bigger than $|x| < k/\sqrt{n}$ so we have a contradiction when $k^2 < b_1/2$.

(B) Given that $|S_n| > k/\sqrt{n}$ the probability that S_M will never return to the interval $|x| < k/\sqrt{n}$ for $M > n$ is at least $a = 4b_2/(4b_2 + k^2$: Assume without loss of generality that $S_n > k/\sqrt{n}$. The sequence S_{n+i} stopped upon re-entering the interval $x < k/2\sqrt{n}$ is a submartingale by (29). When decomposed into a martingale plus an increasing sequence, summing (31) shows that the martingale part has variance never exceeding b_2/n. Then by using the one-sided Tschebysheff estimate $prob(f - \mathbf{E}f > s) \leq \mathrm{Var}(f)/(\mathrm{Var}(f) + s^2)$ and stopping the process if it re-enters the interval$x < k/\sqrt{n}$, we see the probability of the martingale (hence the submartingale) re-entering the interval $x < k/\sqrt{n}$ is bounded above by the constant $4b_2/(4b_2 + k^2)$.

(C) If S_n converges to 0 with non-zero probability, then there is an n which can be chosen arbitrarily large and an event $\mathcal{A} \in \mathcal{F}_n$ for which $prob(S_n \to 0 \,|\, \mathcal{A})$ is arbitrarily close to 1. When it is greater than $1 - a/2$, this contradicts (A) and (B).

Now assume (32) instead of (28) - (31).

(D) Choose $\mathcal{A}$ as in (C) and choose $\epsilon < a/2$ where a is as in (B). Let

$\{X^*_n, S^*_n\}$ be any process that always satisfies (28) - (31) and is coupled to the process $\{X_n, S_n\}$ so that $S_n = S_n^*$ on $\mathcal{B}_1 \cap \cdots \cap \mathcal{B}_n$. Then convergence of S_n to 0 with probability $1 - a/2 + \epsilon$ implies convergence of S_n^* with probability at least $1 - a/2$, contradicting (A) and (B).

For the remainder of this section we will mainly be taking expectations with respect to $\mathcal{F}_n$, so the following notation is convenient.

Definition 5.19 *Let* $\overline{\mathbf{E}}(\cdot)$ *denote* $\mathbf{E}(\cdot \mid \mathcal{F}_n)$.

We must now construct η and $\tau_r, \tau_{r+1}, \ldots$ satisfying (32). For any N_0, let Let $\tau_r = \inf\{k \geq N_0 : Y_k = 1\}$ be the first time after N_0 that the walk visits vertex 1. Let $\tau_{n+1} = \inf\{k > \tau_n : Y_k = 1\}$ be the successive hitting times for vertex 1. We will specify r when necessary. Choose $\mathcal{N}$ small enough so that all coordinates of all points in $\mathcal{N}$ are at least b_3 for some $b_3 > 0$. Let $\tau = \inf\{k \geq N_0 : \vec{v}(k) \notin \mathcal{N}\}$ and replace the old values of τ_i with $\tau \wedge \tau_i$. For ease of notation, let

$$(\Delta\vec{v})_n = \vec{v}(\tau_{n+1}) - \vec{v}(\tau_n).$$

Once again, let $M_{\vec{v}}$ be the markov chain $prob(i \to j) = v_i/N_j$ and let $L(\vec{v})$ be the mean recurrence time for $M_{\vec{v}}$. Let $L_{\max}$ be the supremum of $L(\vec{v})$ over $\mathcal{N}$ and pick $N_0 > 2L_{\max}$. We record a few facts about the times $\tau_r, \tau_{r+1}, \ldots$.

Proposition 5.20 *The distribution of* $\tau_{n+1} - \tau_n$ *has exponential tails. Specifically,*

$$prob(\tau_{n+1} - \tau_n \geq k + 1) < e^{-\alpha k}$$

for some α. *In particular,*

$$\overline{\mathbf{E}}(\tau_{n+1} - \tau_n) < 1/(1 - e^{-\alpha}) \text{ and } \overline{\mathbf{E}}(\tau_{n+1} - \tau_n)^2 < 3e^{-\alpha}/(1 - e^{-\alpha})^2.$$

Proof:

$$\begin{aligned} & prob(\tau_{n+1} - \tau_n \geq k+1) \\ = \; & prob(\tau_{n+1} - \tau_n \geq k-1) \cdot prob(Y_{\tau_n+k} \neq 1 \text{ or } \vec{v}(\tau_n + k) \notin \mathcal{N}) \\ \leq \; & 1 - b_3 \end{aligned}$$

by choice of $\mathcal{N}$. So choose $\alpha = -\ln(1 - b_3)$. □

Proposition 5.21

$$\left| \overline{\mathbf{E}}((\triangle \vec{v})_n) - L(\vec{v}(\tau_n)) \frac{\pi(\vec{v}(\tau_n)) - \vec{v}(\tau_n)}{\tau_n} \right| = O({\tau_n}^2).$$

Proof: Couple the random walk to a Markov chain $\tilde{Y}(\tau_n), \tilde{Y}(\tau_n + 1), \ldots$ with transition matrix $\mathcal{M}_{\vec{v}(\tau_n)}$ starting at time τ_n so that the processes remain coupled as long as possible. Define the twiddled variables $\tilde{\tau}_{n+1} = \inf\{k > \tau_n : \tilde{Y}_k = 1\}$, $\vec{\tilde{S}}$ to be a vector such that $\tilde{S}_i = S_i(\tau_n)$ plus the number of times $\tilde{Y}_k = i$ for $\tau_n < k \leq \tilde{\tau}_{n+1}$, and $\vec{\tilde{v}} = \tilde{S}/\tilde{\tau}_{n+1}$. The Markov property for $\{\tilde{Y}_k\}$ yields the identity

$$\overline{\mathbf{E}}(\vec{\tilde{S}}(\tilde{\tau}_{n+1}) - \vec{S}(\tau_n)) = L(\vec{v}(\tau_n))\pi(\vec{v}(\tau_n)). \tag{33}$$

The probability of the process coming uncoupled before time τ_{n+1} is bounded by

$$\sum_{k \geq 0} prob(\tau_{n+1} - \tau_n > k) k/\tau_n \leq e^{-\alpha}/(1 - e^{-\alpha})^2 \tau_n \tag{34}$$

since the transition probabilities for the two processes differ by at most k/τ_n at time $\tau_n + k$. But for any k,

$$\overline{\mathbf{E}}(\tilde{\tau}_{n+1} - (\tau_n + k) \,|\, \text{ uncoupling occurs at } \tau_n + k) \leq L_{\max}$$

and

$$\overline{\mathbf{E}}(\tau_{n+1} - (\tau_n + k) \,|\, \text{ uncoupling occurs at } \tau_n + k) \leq 1/(1 - e^{-\alpha}).$$

So

$$\begin{aligned}
&\overline{\mathbf{E}}(|\vec{v}(\tau_{n+1}) - \vec{\tilde{v}}(\tilde{\tau}_{n+1})| \mid \text{uncoupling before } \tau_{n+1}) \\
\leq\ & \sup_k \overline{\mathbf{E}}(|\vec{v}(\tau_{n+1}) - \vec{v}(\tau_n)| + |\vec{\tilde{v}}(\tilde{\tau}_{n+1}) - \vec{v}(\tau_n)| \\
& \mid \text{uncoupling occurs at } \tau_n + k) \\
\leq\ & \sup_k (1/\tau_n)(\overline{\mathbf{E}}(\tau_{n+1} - \tau_n - k + 1 \\
& + \tilde{\tau}_{n+1} - \tau_n - k + 1) \mid \text{uncoupling occurs at } \tau_n + k) \\
\leq\ & (1/\tau_n)(L_{\max} + 1/(1 - e^{-\alpha}) + 2).
\end{aligned} \tag{35}$$

Combining (34) and (35) gives

$$\overline{\mathbf{E}}(|\vec{v}(\tau_{n+1}) - \vec{\tilde{v}}(\tilde{\tau}_{n+1})|) = O({\tau_n}^2). \tag{36}$$

Now write

$$\begin{aligned}
& \overline{\mathbf{E}}(\tilde{v}_i(\tilde{\tau}_{n+1}) - v_i(\tau_n)) \\
=\ & \overline{\mathbf{E}}(\tilde{S}_i(\tilde{\tau}_{n+1})/\tilde{\tau}_{n+1} - S_i(\tau_n)/\tau_n) \\
=\ & \overline{\mathbf{E}}\left(\frac{1}{\tilde{\tau}_{n+1}}(\tilde{S}_i(\tau_{n+1}) - S_i(\tau_n) - (\tilde{\tau}_{n+1} - \tau_n)S_i(\tau_n)/\tau_n)\right) \\
=\ & \overline{\mathbf{E}}\left(\frac{1}{\tau_n}(\tilde{S}_i(\tau_{n+1}) - S_i(\tau_n) - (\tilde{\tau}_{n+1} - \tau_n)S_i(\tau_n)/\tau_n)\right) \\
& -\overline{\mathbf{E}}\left(\frac{\tilde{\tau}_{n+1} - \tau_n}{\tilde{\tau}_{n+1}\tau_n}(\tilde{S}_i(\tau_{n+1}) - S_i(\tau_n) - (\tilde{\tau}_{n+1} - \tau_n)S_i(\tau_n)/\tau_n)\right) \\
=\ & \frac{1}{\tau_n}(L(\vec{v}(\tau_n))\pi_i(\vec{v}(\tau_n)) - L(\vec{v}(\tau_n))v_i(\tau_n)) \\
& -\overline{\mathbf{E}}\left(\frac{\tilde{\tau}_{n+1} - \tau_n}{\tilde{\tau}_{n+1}} \frac{(\tilde{S}_i(\tau_{n+1}) - S_i(\tau_n) - (\tilde{\tau}_{n+1} - \tau_n)S_i(\tau_n)/\tau_n)}{\tau_n}\right)
\end{aligned}$$

according to (33). Combining this with (36) and using the identity $\overline{\mathbf{E}}(|fg|) \leq \overline{\mathbf{E}}(f^2 + g^2)$ shows that the left side of proposition 5.21 is at most

$$O({\tau_n}^2) + (d/{\tau_n}^2)(\overline{\mathbf{E}}(L(\vec{v}(\tau_n)))^2 + \overline{\mathbf{E}}(\tilde{\tau}_{n+1} - \tau_n)^2) = O({\tau_n}^2)$$

by proposition 5.20. □

Proposition 5.22 *For any $\epsilon > 0, r, N_0$, we can pick constants, c_1 and c_2 such that*

$$prob(c_1 n \leq \tau_n \leq c_2 n \text{ for all } n \geq r \text{ such that } \vec{v}(\tau_n) \in \mathcal{N}) > 1 - \epsilon.$$

Proof: This is a consequence of the fact that the variables $\tau_{n+1} - \tau_n$ are stochastically bounded above by independent geometric random variables as in proposition 5.20, and below by 1 as long as $\vec{v}(\tau_n) \in \mathcal{N}$. □.

Proposition 5.23 *For any $\epsilon > 0, \gamma < 1$, we can pick N_0, r large enough so that*

$$prob(\tau_{n+1} - \tau_n \geq n^{1-\gamma} \text{ for some } n \geq r) < \epsilon.$$

Proof: This follows from proposition 5.20 and the fact that $\sum e^{-\alpha(n^{1-\gamma})} < \infty$. □

The final step is to construct η. The idea behind η is simple. The linear approximation, T, to $\pi - I$ has an eigenvector with a positive eigenvalue λ and an invariant subspace W_1 not containing this eigenvector. Since the expected change from $\vec{v}(\tau_n)$ to $\vec{v}(\tau_{n+1})$ is in the direction $(\pi - I)\vec{v}(\tau_n)$ by proposition 5.21, the distance between $\vec{v}$ and W_1 should increase by a factor of $1 + \lambda$ on the average. So if $\pi - I$ were linear and $\eta(\vec{v})$ was defined to be $\vec{v} \cdot \vec{\theta}$ for some $\vec{\theta} \perp W_1$, then the definition (27) should make $\overline{\mathbf{E}} X_{n+1}$ the same sign as S_n as long as the latter is far enough from 0 so that the error in approximating $\overline{\mathbf{E}} S_{n+1}$ using proposition 5.21 does not cause it to change sign. Then (29) will be true, and the rest of (32) would follow easily from the preceding propositions.

Unfortunately $\pi - I$ is not linear, so there is no plane such that $\pi - I$ always points away from it. However, if $\pi - I$ is non-singular, there is always a surface of codimension 1 which is an invariant manifold for the flow (3) and which the flow moves away from when perturbed. We can use this to define η but our approximations will only work when the

surface is twice continuously differentiable. To get this differentiability it is sufficient to impose the extra hypothesis on the eigenvalues that appears in theorem 5.12, although such a condition hardly seems necessary from a stochastic viewpoint. We call upon the following standard result from the theory of ODE's.

Theorem 5.24 (Sternberg 1958) *Suppose T is a real $d \times d$ non-singular matrix with eigenvalues linearly independent over the rationals and none being purely imaginary. Suppose*

$$F(\vec{v}) = T\vec{v} + \chi(\vec{v})$$

is $\mathbf{C}^\infty$ in a neighborhood of $\vec{0}$, with $\chi(\vec{0}) = \vec{0}$ and its differential $\partial_{\vec{v}}\chi(\vec{0}) \equiv 0$. Let $G(\vec{v}, t)$ denote the (necessarily smooth and unique) map defined on a neighborhood of $(\vec{0}, 0)$ such that $d/dt(G(\vec{v}, t)) = F(G(\vec{v}, t))$ and $G(\vec{v}, 0) = \vec{v}$. Let $F^t(\vec{v}) = G(\vec{v}, t)$. Then there is a $\mathbf{C}^\infty$ diffeomorphism Φ from a neighborhood of $\vec{0}$ to a neighborhood of $\vec{0}$ such that

$$\begin{aligned} \Phi(\vec{0}) &= \vec{0} \\ \Phi F^t \Phi^{-1} &= e^{tT}. \end{aligned} \tag{37}$$

Proof: See [Ha2] theorem 12.1 and exercise 12.1, page 257. □

Apply this theorem with $F(\vec{v}) = (\pi - I)(\vec{v} + \vec{p})$. Differentiating (37) with respect to t at $t = 0$, we see that

$$\partial_{\vec{v}}\Phi(\pi - I)(\vec{v} + \vec{p}) = T\Phi(\vec{v}) \tag{38}$$

Now let $\eta(\vec{v}) = \Phi(\vec{v} - \vec{p}) \cdot \vec{\theta}$ where $\vec{\theta}$ is normal to the invariant subspace W_1. Recall that this means

$$\begin{aligned} S_n &= \Phi(\vec{v}(\tau_n)) \cdot \vec{\theta} \\ X_n &= S_n - S_{n-1}. \end{aligned}$$

We must establish (32). Firstly, note that Φ and Φ^{-1} are bounded in the operator norm on a neighborhood of $\vec{0}$, i.e. for some $c_1, c_2 > 0$ we have

$$c_1|\vec{v} - \vec{w}| \leq |\Phi(\vec{v}) - \Phi(\vec{w})| \leq c_2|\vec{v} - \vec{w}| \tag{39}$$

for $\vec{v}, \vec{w} \in \mathcal{N}$ where $\mathcal{N}$ is sufficiently small. Then for any $\epsilon > 0$ we can get (30) from propostition 5.23 and we can get (31) from propositions 5.20 and 5.22 with probability at least $1 - \epsilon$. To prove (29), let $\epsilon > 0$ be arbitrary and calculate

$$\begin{aligned}
\overline{\mathbf{E}}(X_{n+1}) &= \overline{\mathbf{E}}(S_{n+1}) - S_n \\
&= \overline{\mathbf{E}}(\eta(\vec{v}(\tau_n) + (\Delta\vec{v})_n)) - S_n \\
&= \vec{\theta} \cdot \overline{\mathbf{E}}(\Phi(\vec{v}(\tau_n) + (\Delta\vec{v})_n)) - S_n \\
&= \vec{\theta} \cdot \overline{\mathbf{E}}(\Phi(\vec{v}(\tau_n)) + \partial_{\vec{v}}\Phi((\Delta\vec{v})_n) + O(|(\Delta\vec{v})_n|^2)) - S_n \\
&= \vec{\theta} \cdot \partial_{\vec{v}}\Phi((\Delta\vec{v})_n) + O(\overline{\mathbf{E}}|(\Delta\vec{v})_n|^2) \\
&= \vec{\theta} \cdot \partial_{\vec{v}}\Phi(L(\vec{v}(\tau_n))(\pi - I)\vec{v}(\tau_n)/\tau_n + O(1/n^2) + O(\overline{\mathbf{E}}|(\Delta\vec{v})_n|^2) \\
&\quad \text{with probability } \geq 1 - \epsilon, \text{ by propositions 5.21 and 5.22} \\
&= L(\vec{v}(\tau_n))\vec{\theta} \cdot T\Phi(\vec{v}(\tau_n) - \vec{p})/\tau_n + O(1/n^2) \\
&\quad \text{by (38) and proposition 5.20} \\
&= \lambda S_n/tau_n + O(1/n^2).
\end{aligned} \tag{40}$$

According to proposition 5.22 there are constants c_1 and c_2 for which $c_1 n \leq \tau_n \leq c_2 n$, hence there is a constant c such that for $S_n > c/n$ the first term of (40) dominates. Thus (29) is true with probability at least $1 - \epsilon$.

Finally, to show (28), note that there is a set of d events $\mathcal{A}_1, \ldots, \mathcal{A}_d$, each corresponding to a single sequence of values $Y_{\tau_n+1}, \ldots, Y_{\tau_n+r-1}, Y_{\tau_{n+1}} = 1$, such that $v_j(\tau_{n+1}) > v_j \Leftrightarrow i = j$ on $\mathcal{A}_i$. Then the probability of each $\mathcal{A}_d$ is bounded below by a constant times ${\tau_n}^{-1}$, which together with equation (39) and proposition 5.22 shows that r and N_0 can be picked so $\overline{\mathbf{E}}(X_{n+1})^2$ is at least a constant times n^{-2} with probability arbitrarily close to 1. When $|S_n > c/n|$, equation (28) now follows from (40). When $|S_n \leq c/n|$,

the second term in (40) dominates, but then $S_n\overline{\mathbf{E}}(X_{n+1})$ is $O(1/n^3)$ so for n sufficiently large, (28) is true.

5.6 Examples

EXAMPLE 1: Consider the VRRW on a complete graph without loops on d vertices. So $\mathbf{R}_{ij} = 1 - \delta_{ij}$ for $1 \leq i, j \leq d$. When $d = 3$ this is just the triangle G_1 of section 5.1. The critical points are the centroids of the faces of $\triangle$, and it is easy to see that the centroids of all proper faces are linear non-maxima. For example, if $\vec{p} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \ldots, 0)$ then $N_1(\vec{p}) = N_2(\vec{p}) = N_3(\vec{p}) = 2/3$ but $N_i(\vec{p}) = 1$ for $i > 3$ so $\vec{p}$ is a linear non-maximum by the criterion (19). It follows from theorems 5.8 and 5.11 that $\vec{v}(n) \to (\frac{1}{d}, \ldots, \frac{1}{d})$. In particular, VRRW on the triangle has $\vec{v}(n) \to (1/3, 1/3, 1/3)$.

EXAMPLE 2: Consider the VRRW on a $d-$sided polygon for $d \geq 5$. So

$$\mathbf{R}_{i,j} = \begin{cases} 1 \text{ if } i - j = 1 \bmod d \\ 0 \text{ otherwise.} \end{cases}$$

When $d = 5$ this is the graph G_3 of section 5.1. As previously mentioned, there are too many zero entries and duplications of eigenvalues to permit application of theorems 5.11 and 5.12. Nevertheless, in the spirit of corollary 5.13, we can say things about "almost all" small perturbations of $\mathbf{R}$.

The critical set $\mathcal{C}$ consists of the isolated point $(1/d, \ldots, 1/d)$ together with points on the boundary of $\triangle$ of the form $(0, \ldots, 0, x, 1/2, 1/2-x, 0, \ldots, 0)$ or convex combinations of such points. Any small perturbation of $\mathbf{R}$ has an isolated interior critical point and other critical points near the boundary of $\triangle$. Since $\mathbf{R}$ is a circulant matrix, i.e. the $i,j-$entry depends only on $i - j \bmod d$, the eigenvalues of $\mathbf{R}$ are easily calculated (see example 5 below). They are in fact $\{2\text{Re}(e^{2i\pi k/d} : 0 \leq k \leq d-1\}$. For $d \geq 5$ this list contains more than one positive value. Applying corollary 5.13, there is a neighborhood

of $(1/d, \ldots, 1/d)$ such that no sufficiently small generic perturbation of $\mathbf{R}$ gives rise to a VRRW that can converge in that neighborhood. I strongly believe this result to be true for $\mathbf{R}$ itself. Computer simulation bears out the guess that VRRW on a polygon of more than 4 sides always concentrates asymptotically on a set of three consecutive vertices and never approaches the uniform distribution.

EXAMPLE 3: Consider the VRRW on the square. So

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

This example has nothing to do with the theorems of this chapter, but it completes the story on polygons. Let the vertices be numbered cyclically $1, 2, 3$ and 4. Since the graph is bipartite, the walk must alternate between being on vertices 1 and 3 and being on vertices 2 and 4. Furthermore the choices made between vertices 1 and 3 at each second move are completely independent of the choices between vertices 2 and 4 on the intervening moves; the probability of moving to vertex 1 is always $S_1(n)/(S_1(n)+S_3(n))$ regardless of whether the walk is at vertex 2 or 4 at move n. Clearly, the sequence of choices between vertices 1 and 3 behaves like a sequence of draws from a two-color Pólya urn. Applying the basic result on Pólya's urn (theorem 2.1) with $R = B = \Delta = 1$, the proportion of times vertex 1 is chosen over vertex 3 will approach a limit that is uniform over $(0, 1)$. The same is true for the sequence of choices between vertices 2 and 4. Then $\vec{v}(n)$ converges to a random vector $\vec{v}$ whose distribution is uniform over the affine square $\{\vec{v} : v_1 + v_3 = v_2 + v_4 = 1/2\}$.

EXAMPLE 4: When $d = 2$ the VRRW can be completely described for almost every matrix $\mathbf{R}$. Let

$$\mathbf{R} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

and assume $b \neq 0$ to avoid triviality. Then Δ is the line segment joining $(1,0)$ and $(0,1)$. The endpoints are always critical points, and in addition there is a critical point at $\left(\frac{c-b}{a+c-2b}, \frac{a-b}{a+c-2b}\right)$ whenever $(a-b)(c-b) > 0$. We consider three cases, depending on the sign of $(a-b)(c-b)$. (All convergence is taken to be almost sure convergence.)

Case 1: $(a-b)(c-b) < 0$.
Assume without loss of generality that $a < b < c$. Then H has a linear non-maximum at $(1,0)$ so by theorems 5.8 and 5.11, $\vec{v}(n)$ must converge to the point $(0,1)$ as $n \to \infty$.

Case 2: $(a-b)(c-b) = 0$.
Assume without loss of generality that $b = c$.

Case 2a: $a < b = c$. Then $(1,0)$ is a linear non-maximum again and $\vec{v}(n) \to (0,1)$.

Case 2b: $a > b = c$. Now $(0,1)$ is a global minimum for H but it is not a linear non-maximum because the derivative of H vanishes here. Theorem 5.8 tells us that $\vec{v} \to \vec{p}$ where $\vec{p}$ is one of hte two endpoints of the interval, and while the author suspects it must always be $(1,0)$, we have no proof.

Case 2c: $a = b = c$. Now the critical set $\mathcal{C}$ is the whole simplex. Our theorems tell us nothing a priori, but by viewing the process as a simple Pólya urn process, we can see that $\vec{v}(n) \to \vec{p}$ where $\vec{p}$ is a random variable whose distribution is uniform on the simplex.

Case 3: $(a-b)(c-b) > 0$.

Case 3a: $a < b$ and $c < b$. Then both $(0,1)$ and $(1,0)$ are linear non-maxima, but there is another critical point $\left(\frac{c-b}{a+c-2b}, \frac{a-b}{a+c-2b}\right)$ in the interior

of the simplex to which $\vec{v}(n)$ must converge.

Case 3b: $a > b$ and $c > b$. Now it is possible for $\vec{v}(n)$ to converge to either endpoint. There is a critical point $\left(\frac{c-b}{a+c-2b}, \frac{a-b}{a+c-2b}\right)$ in the interior as well, which is a minimum for H, since both eigenvalues for $\mathbf{R}$ are positive. If the eigenvalues of $diag(\vec{p})\mathbf{R}$ are incommeasurable, then theorem 5.12 says that $prob(\vec{v}(n) \to \left(\frac{c-b}{a+c-2b}, \frac{a-b}{a+c-2b}\right)) = 0$. When the eigenvalues are commeasurable, theorem 5.12 does not apply. We can however define η easily enough in this case without resorting to Sternberg's theorem. For example, let η be projection onto the first coordinate, centered so that $\eta\left(\frac{c-b}{a+c-2b}, \frac{a-b}{a+c-2b}\right) = 0$. It follows that $\vec{v}(n)$ still cannot converge to $\left(\frac{c-b}{a+c-2b}, \frac{a-b}{a+c-2b}\right)$.

EXAMPLE 5: Let G be a finite abelian group and let $T \subseteq G$ be a set of generators and be closed under inverses. Let $\mathbf{R}$ be the incidence matrix for the Cayley graph associated with these generators, so the rows and columns of $\mathbf{R}$ are indexed by G, and $\mathbf{R}_{gh} = 1$ when $gh^{-1} \in T$ and 0 otherwise. Example 1 is a special case of this where G is any abelian group of order d and T is G minus the identity. Example 2 is the special case where G is cyclic and $T = \{1, -1\}$.

Calculating the eigenvalues of $\mathbf{R}$ is easy. For any character χ on G, let $\chi(T)$ denote $\sum_{g \in T} \chi(g)$. Then as χ ranges over all characters, $\chi(T)$ ranges over all eigenvalues of $\mathbf{R}$. The point $\vec{p} = (\frac{1}{d}, \ldots, \frac{1}{d})$ is always a critical point for H and when all the values of $\chi(T)$ are non-zero, it is an isolated critical point. If furthermore, all values of $\chi(T)$ are negative except when χ is the trivial character, then H has a global maximum at $\vec{p}$ and it is not hard to see that all other critical points are linear non-maxima, so $\vec{v} \to \vec{p}$. This is the case in example 1. In the remaining cases we would like to apply theorem 5.12 but cannot because the eigenvalues will always come in pairs and thus be (very) linearly dependent over the rationals. We must settle for the following observation: when $\chi(T) > 0$ for some non-trivial character then there are arbitrarily

small perturbations of $\mathbf{R}$ for which theorem 5.12 does apply and for which process $\vec{v}(n)$ cannot converge to $\vec{p}$. We therefore suspect that $\vec{v}(n)$ does not converge to $\vec{p}$ in the unperturbed case.

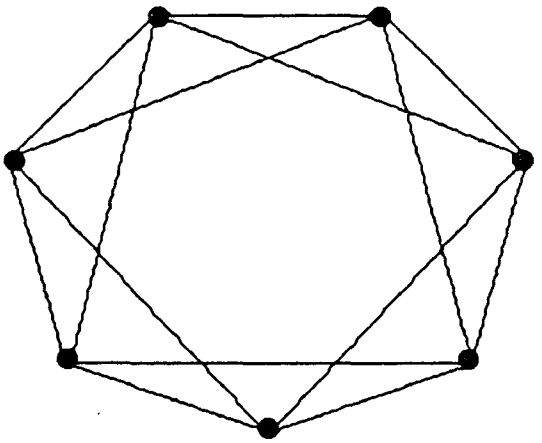
The following somewhat longwinded analysis illustrates the calculations involving characters and also shows how to get further information from theorem 5.11 by looking at all subgraphs of the original graph. Let G be the integers mod 7 and let $T = \{1, 2, 5, 6\}$. So the VRRW takes places on the Cayley graph G_4 of figure 7 that is just a heptagon plus all its short diagonals. (By applying the automorphisms $1 \mapsto 2$ or $1 \mapsto 3$, results for this case are also seen to apply to the case $T = \{2, 3, 4, 5\}$ or $T = \{1, 3, 4, 6\}$, or pictorially, a heptagon with any two of the three types of edges present.)

Since $\mathbf{R}$ is nonsingular, the only critical point in the interior of Δ is $(1/7, \ldots, 1/7)$. To see if this is stable, compute the eigenvalues of $\mathbf{R}$. The characters of G are the homomorphisms taking a generator of G to $e^{2i\pi k/7}$ for k $= 0, 1, \ldots, 6$. So the eigenvalues of $\mathbf{R}$ are

$$\omega + \omega^2 + \omega^5 + \omega^6$$

for $\omega = e^{2i\pi k/7}$. For $k = 0$ the eigenvalue is always positive. For $k = 2, 3, 4, 5$ the eigenvalues are negative, but for $k = 1$ and 6 they are positive. Then the critical point is unstable and the VRRW does not converge near $(1/7, \ldots, 1/7)$ for generic small perturbations of $\mathbf{R}$.

To see where it does converge, look for critical points on the faces of Δ. A vector on a proper face of Δ is supported on a subgraph, and there is a correspondence between all faces of Δ and subgraphs of G_4, with $face(\vec{v})$ corresponding to the subgraph on the support of $\vec{v}$ (those vertices j for which $v_j > 0$). The highest dimensional proper faces are subgraphs on six vertices. These are all isomorphic, so consider any of the 6-vertex subgraphs and let the vertices be A, B, C, D, E and F in cyclic order, where the missing vertex is between F and A. If there is a critical point on the interior of a maximal face for the VRRW on G_4, then there is a critical point for this subgraph



$$R = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

FIGURE 7

that is interior to $\triangle$. By proposition 5.4, all the values $N_A, \ldots, N_F$ must be equal at the critical point. Then the following six quantities are equal:

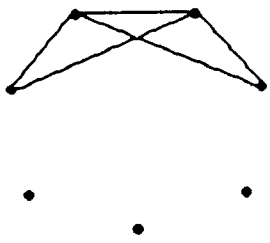
$$\begin{array}{ccccccccc} v_B & + & v_C & + & v_F & & & & \\ v_A & + & v_C & + & v_D & & & & \\ v_A & + & v_B & + & v_D & + & v_E & & \\ v_B & + & v_C & + & v_E & + & v_F & & \\ v_C & + & v_D & + & v_F & & & & \\ v_A & + & v_D & + & v_E & & & & . \end{array}$$

Furthermore, if $p, q, r, s, t, u)$ is a critical point, then by symmetry so is (u, t, s, r, q, p) and hence by linearity so is $(p+u, q+t, r+s, r+s, q+t, p+u)/2$. To determine that there are no interior critical points, it suffices therefore to look only at points with $v_A = v_F, v_B = v_E$ and $v_C = v_D$. Then the equality of the first and third of the six quantities above implies

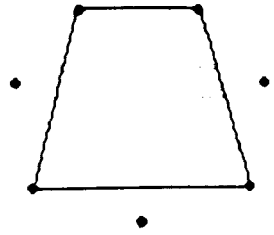
$$v_B + v_C + v_A = v_A + v_B + v_C + v_B$$

which is impossible on the interior of the simplex as v_B cannot be zero on the interior of $\triangle$. Therefore there are no critical points interior to the five-dimensional faces of $\triangle$.

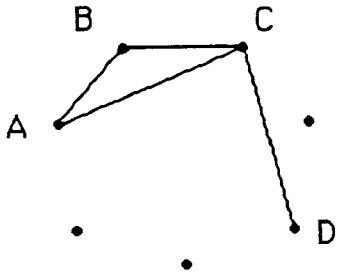
To check for critical points interior to four-dimensional faces, there are three types of isomorphism classes of subgraphs on five vertices to examine. The same method as above, shows that none of these faces actually contains a critical point. When we get to subgraphs of size four, it turns out that two of the types do contain critical points (types 1 and 2 in figure 8), while the two types at the bottom of figure 8 do not. The second type however gives rise to a linear non-maximum (see definition 5.10) because the value of N_i for the four vertices present is $1/2$, while there must be a vertex of the original graph adjacent to the two vertices with the highest weights which will then have $N > 1/2$. Then theorem 5.11 rules out convergence to a point of type 2 except for the point $x = y = 1/4$. Every subgraph on three vertices contains a critical point, but

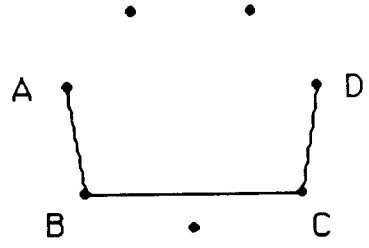


TYPE 1



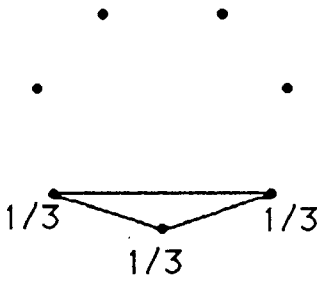
TYPE 2



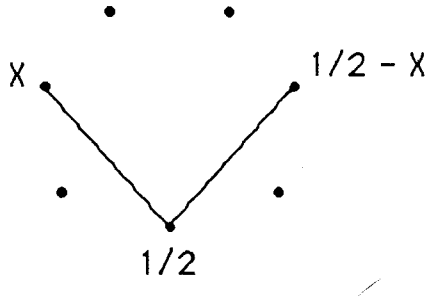


$$N_B = V_A + V_C > V_C = N_D$$

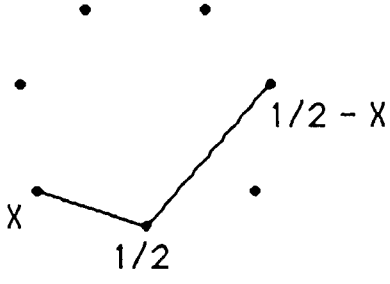
FIGURE 8



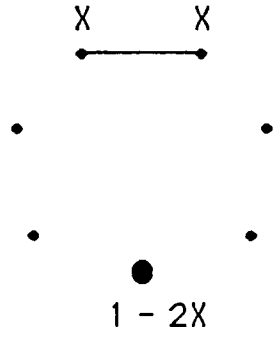
TYPE 3



TYPE 4



TYPE 5



TYPE 6

FIGURE 9

all of them are linear non-maxima except for type 3 in figure 9. Similarly, all 2-vertex and 1-vertex subgraphs have critical points but they are all linear non-maxima. The only candidate is therefore the critical point of type 1, or of type 3, which is just type 1 with $x = 0$.

For a small perturbation of $\mathbf{R}$, the one parameter family of critical points of type 1 will collapse to a single point. As the perturbation gets smaller, this point must approach the one parameter family though it need not have a limit. Thus VRRW on small perturbations of $\mathbf{R}$ will converge somewhere near a point of type 1. Again, I believe that VRRW on $\mathbf{R}$ itself converges to a point of type 1 almost surely.