Random Walks with "Back Buttons"

Ronald Fagin* Anna R. Karlin† Jon Kleinberg‡ Prabhakar Raghavan*
Sridhar Rajagopalan* Ronitt Rubinfeld§ Madhu Sudan¶ Andrew Tomkins*

Abstract

We introduce *backoff processes*, an idealized stochastic model of browsing on the world-wide web, which incorporates both hyperlink traversals and use of the "back button." With some probability the next state is generated by a distribution over out-edges from the current state, as in a traditional Markov chain. With the remaining probability, however, the next state is generated by clicking on the back button, and returning to the state from which the current state was entered. Repeated clicks on the back button require access to increasingly distant history.

We show that this process has fascinating similarities to and differences from Markov chains. In particular, we prove that like Markov chains, backoff processes always have a limiting distribution, and we give algorithms to compute this distribution. Unlike Markov chains, the limiting distribution may depend on the initial state.

*IBM Almaden Research Center, Department K53/B2, 650 Harry Road, San Jose, CA 95120. {fagin,pragh,sridhar,tomkins}@almaden.ibm.com

†Department of Computer Science, Box 352350, University of Washington, Seattle, WA 98195. karlin@cs.washington.edu

‡Department of Computer Science, Cornell University, Ithaca, NY 14853. kleinber@cs.cornell.edu

§NECI, 4 Independence Way, Princeton, NJ 08540. On leave from Cornell University. ronitt@research.nj.nec.com

¶MIT Laboratory for Computer Science, 545 Technology Square NE43-307, Cambridge, MA 02139. madhu@mit.edu

1 Introduction

Consider a modification of a Markov chain in which at each step, with some probability, we *undo* the last forward transition of the chain. For intuition, the reader may wish to think of a user using a browser on the world-wide web where he is following a Markov chain on the pages of the web, and occasionally hitting the "back button". We model such phenomena by discrete-time stochastic processes of the following form: we are given a Markov chain M on a set $V = \{1, 2, \ldots, n\}$ of states, together with an n-dimensional vector $\vec{\alpha}$ of *backoff probabilities*. The process evolves as follows: at each time step $t = 0, 1, 2, \ldots$, the process is in a state $X_t \in V$, and in addition has a *history* H_t, which is a stack whose items are states from V. Let $\text{top}(H)$ denote the top of the stack H. At $t = 0$ the process starts at some state $X_0 \in V$, with the history H_0 containing only the single element X_0. At each subsequent step the process makes either a *forward step* or a *backward step*, by the following rules: (i) if H_t consists of the singleton X_0 it makes a forward step; (ii) otherwise, with probability $\alpha_{\text{top}(H_t)}$ it makes a backward step, and with probability $1 - \alpha_{\text{top}(H_t)}$ it makes a forward step. The forward and backward steps at time t are as follows:

1. In a *forward step*, X_t is distributed according to the successor state of X_{t-1} under M; the state X_t is then pushed onto the history stack H_{t-1} to create H_t.

2. In a *backward step*, the process pops $\text{top}(H_{t-1})$ from H_{t-1} to create H_t; it then moves to $\text{top}(H_t)$ (i.e., the new state $X_t = \text{top}(H_t)$.)[1]

Under what conditions do such processes have limiting distributions, and how do they differ from traditional Markov chains? We focus in this paper on the time-averaged limit distribution, usually called the "Cesaro limit distribution".[2]

Motivation. Our work is broadly motivated by user modeling for scenarios in which a user with an "undo" capability performs a sequence of actions. A simple concrete setting is that of browsing on the world-wide web. We view the pages of the web as states in a Markov chain, with the transition probabilities denoting the distribution over new pages to which the user can move forward, and the backoff vector denoting for each state the probability that a user enters the state and elects to click the browser's back button rather than continuing to browse forward from that state.

A number of research projects [1, 7, 9] have designed and implemented web intermediaries and learning agents that build simple user models, and used them to personalize the user experience. On the commercial side, user models are exploited to better target advertising on the web based on a user's browsing patterns; see [2] and references therein for theoretical results on these and related problems. Understanding more sophisticated models such as ours is interesting in its own right, but could also lead to better user modeling.

[1]Note that the condition $X_t = \text{top}(H_t)$ holds for all t, independent of whether the step is a forward step or backward step.

[2]The *Cesaro limit* of a sequence $a_0, a_1, \ldots$ is $\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t-1} a_\tau$, if the limit exists. For example, the sequence 0,1,0,1,... has Cesaro limit 1/2. The *Cesaro limit distribution* is $\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t-1} \Pr[X_t = i]$, if the limit exists. By contrast, the *stationary distribution* is $\lim_{t\to\infty} \Pr[X_t = i]$, if the limit exists. Of course, a stationary distribution is always a Cesaro limit distribution. We shall sometimes refer simply to either a stationary distribution or a Cesaro limit distribution as a *limiting distribution*.

Overview of Results

For the remainder of this paper we assume a finite number of states. We assume also that the backoff process is *irreducible* (i.e., it is possible, with positive probability, to eventually reach each state from each other state). We now give the reader a preview of some interesting and arguably unexpected phenomena that emerge in such "back-button" random walks, as the backoff vector $\vec{\alpha}$ varies on a fixed Markov chain M. Our primary focus is on the Cesaro limit distribution.

Intuitively, if the history stack H_t grows unboundedly with time, then the process "forgets" the start state X_0 (as happens in a traditional Markov process, where $\vec{\alpha}$ is identically zero). On the other hand, if the elements of $\vec{\alpha}$ are all very close to 1, the reader may envision the process repeatedly "falling back" to the start state X_0, so that H_t does not tend to grow unboundedly. What happens between these extremes?

One of our main results is that there is always a Cesaro limit distribution, although there may not be a stationary distribution. Consider first the case when all entries of $\vec{\alpha}$ are equal, so that there is a single backoff probability α that is independent of the state. In this case we give a remarkably simple characterization of the history and the limiting distribution provided $\alpha < 1/2$: the history grows unboundedly with time, and the limiting distribution of the process converges to that of the underlying Markov chain M.

On the other hand, if $\alpha > 1/2$ then the process returns to the start state X_0 infinitely often, the expected history length is finite, and the limiting distribution differs in general from that of M, and depends on the start state X_0. Thus, unlike ergodic Markov chains, the limiting distribution depends on the starting state.

More generally, consider starting the backoff process in a probability distribution over the states of M; then the limiting distribution depends on this initial distribution. As the initial distribution varies over the unit simplex, the set of limiting distributions forms a simplex. As α converges to $1/2$ from above, these simplices converge to a single point, which is the limiting distribution of the underlying Markov chain.

The transition case $\alpha = 1/2$ is fascinating: the process returns to the start state infinitely often, but the history grows with time and the distribution of the process reaches the stationary distribution of M. These results are described in Section 3.

We have distinguished three cases: $\alpha < 1/2$, $\alpha = 1/2$, and $\alpha > 1/2$. In Section 4, we show that these three cases can be generalized to backoff probabilities that vary from state to state. The generalization depends on whether a certain infinite Markov process (whose states correspond to possible histories) is transient, null, or ergodic respectively (see Section 4 for definitions). It is intuitively clear in the constant α case, for example, that when $\alpha < 1/2$, the history will grow unboundedly. But what happens when some states have backoff probabilities greater than $1/2$ and others have backoff probabilities less than $1/2$? When does the history grow, and how does the limiting distribution depend on M and $\vec{\alpha}$? Even when all the backoff probabilities are less than $1/2$, why should there be a limiting distribution?

We resolve these questions by showing that there exists a potential function of the history that is expected to grow in the transient case (where the history grows unboundedly), is expected to shrink in the ergodic case (where the expected size of the history stack remains bounded), and is expected to remain constant if the process is null. The potential function is a bounded difference martingale, which allows us to use martingale tail inequalities to prove these equivalences. Somewhat surprisingly, we can use this relatively simple characterization of the backoff process to obtain an efficient algorithm to decide, given M and α, whether or not the given process is transient, null or ergodic. We show that in all cases the process attains a Cesaro limit distribution (though the proofs are quite different for the different cases). We also give algorithms to compute the limiting probabilities. If the process is either ergodic or null then the limiting probabilities are computed exactly by solving certain systems of linear inequalities. However, if the process is transient, then the limiting probabilities need not be rational numbers, even if all entries of M and $\vec{\alpha}$ are rational. We show that in this case, the limiting probabilities can be obtained by solving a linear system, where the entries of the linear system are themselves the solution to a semidefinite program. This gives us an algorithm to approximate the limiting probability vector.

2 Definitions and notation

We use $(M, \vec{\alpha}, i)$ to denote the backoff process on an underlying Markov chain M, with backoff vector $\vec{\alpha}$, starting from state i. This process is an (infinite) Markov chain on the space of all histories. Formally, a *history stack* (which we may refer to as simply a *history*) $\bar{\sigma}$ is a sequence $\langle \sigma_0, \sigma_1, \ldots, \sigma_l \rangle$ of states of V, for $l \geq 0$. For a history $\bar{\sigma} = \langle \sigma_0, \sigma_1, \ldots, \sigma_l \rangle$, its *length*, denoted $\ell(\bar{\sigma})$, is l (since the initial state σ_0 is special, we do not count it in the length). If $\ell(\bar{\sigma}) = 0$, then we say that it is an *initial* history. For a history $\bar{\sigma} = \langle \sigma_0, \sigma_1, \ldots, \sigma_l \rangle$, then its *top*, denoted $\mathrm{top}(\bar{\sigma})$, is σ_l. We also associate the standard stack operations pop and push with histories. For a history $\bar{\sigma} = \langle \sigma_0, \sigma_1, \ldots, \sigma_l \rangle$ and $j \in \{1, \ldots, n\}$, we have $\mathsf{pop}(\bar{\sigma}) = \langle \sigma_0, \sigma_1, \ldots, \sigma_{l-1} \rangle$, and $\mathsf{push}(\bar{\sigma}, j) = \langle \sigma_0, \sigma_1, \ldots, \sigma_l, j \rangle$. We let $\mathcal{S}$ denote the space of all finite attainable histories.

For a Markov chain M, backoff vector $\vec{\alpha}$, and history $\bar{\sigma}$ with $\mathrm{top}(\bar{\sigma}) = j$, define the successor (or next state) $\mathrm{succ}(\bar{\sigma})$ to take on values from $\mathcal{S}$ with the following distribution:

$$\mathrm{succ}(\bar{\sigma}) = \begin{cases} \mathsf{pop}(\bar{\sigma}) & \text{with probability } \alpha_j & \text{if } \ell(\bar{\sigma}) \geq 1 \\ \mathsf{push}(\bar{\sigma}, k) & \text{with probability } (1 - \alpha_j) M_{jk} & \text{if } \ell(\bar{\sigma}) \geq 1 \\ \mathsf{push}(\bar{\sigma}, k) & \text{with probability } M_{jk} & \text{if } \ell(\bar{\sigma}) = 0 \end{cases}$$

For a Markov chain M, backoff vector $\vec{\alpha}$ and state $i \in \{1, \ldots, n\}$, the $(M, \vec{\alpha}, i)$*-Markov chain* is the sequence $\langle H_0, H_1, H_2, \ldots \rangle$ taking values from the set $\mathcal{S}$ of histories, with $H_0 = \langle i \rangle$ and H_{t+1} distributed as $\mathrm{succ}(H_t)$. We refer to the sequence $\langle X_0, X_1, X_2, \ldots \rangle$, with $X_t = \mathrm{top}(H_t)$ as the $(M, \vec{\alpha}, i)$*-backoff process*. Several properties of the $(M, \vec{\alpha}, i)$-backoff process are actually independent of the start state i, and to stress this aspect we will sometimes use simply the term "$(M, \vec{\alpha})$-backoff process".

Note that the $(M, \vec{\alpha}, i)$-backoff process does not completely give the $(M, \vec{\alpha}, i)$-Markov chain, because it does not specify whether each step results from a "forward" or "backward" operation. To complete the correspondence we define an *auxiliary sequence*: Let $S_1, \ldots, S_t, \ldots$ be the sequence with S_t taking on values from the set $\{F, B\}$, with $S_t = F$ if $\ell(H_t) = \ell(H_{t-1}) + 1$ and $S_t = B$ if $\ell(H_t) = \ell(H_{t-1}) - 1$. (Intuitively, F stands for "forward" and B for "backward".) Notice that sequence $X_0, \ldots, X_t, \ldots$ together with the sequence $S_1, \ldots, S_t, \ldots$ does completely specify the sequence $H_0, \ldots, H_t, \ldots$.

We study the distribution of the states X_t as the backoff process evolves over time. We shall show that there is always a Cesaro limit distribution (although there is not necessarily a stationary distribution). We shall also study the question of efficiently computing the Cesaro limit distribution.

For simplicity, throughout this paper, we shall restrict our attention to cases where both the $(M, \vec{\alpha})$-backoff process and the underlying Markov chain M are irreducible and aperiodic. In particular, M has a stationary distribution, and not just a Cesaro limit distribution.

3 Constant α

The case in which the backoff probability takes the same value α for every state has a very clean characterization, and it will give us insight into some of the arguments to come.

We fix a specific $(M, \vec{\alpha}, i)$-backoff process throughout this section. Suppose we generate a sequence $X_0, X_1, \ldots, X_t, \ldots$ of steps together with an auxiliary sequence $S_1, \ldots, S_t, \ldots$. To begin with, we wish to view this sequence of steps as being "equivalent" (in a sense) to one in which only forward steps are taken. In this way, we can relate the behavior of the $(M, \vec{\alpha}, i)$-backoff process to that of the underlying (finite) Markov process M beginning in state i, which we understand much more accurately. We write $q_t(j)$ to denote the probability that M, starting in state i, is in state j after t steps.

When the backoff probability takes the same value α for every state, we have the following basic relation between these two processes.

Theorem 3.1 *For given natural numbers λ and t, and a state j, we have* $\Pr[X_t = j \mid \ell(H_t) = \lambda] = q_\lambda(j)$.

The proof of this theorem is given in Appendix B.

In addition to the sequences $\{X_t\}$ and $\{S_t\}$, consider the sequence $\{Y_t : t \geq 0\}$, where Y_t is the history length $\ell(H_t)$. Now Y_t is simply the position after t steps of a random walk on the natural numbers, with a reflecting barrier at 0, in which the probability of moving left is α and the probability of moving right is $1 - \alpha$. This correspondence will be crucial for our analysis.

In terms of these notions, we mention one additional technical lemma. Its proof follows simply by conditioning on the value of Y_t and applying Theorem 3.1.

Lemma 3.2 *For all natural numbers t and states j, we have* $\Pr[X_t = j] = \sum_r q_r(j) \cdot \Pr[Y_t = r]$.

We are now ready to consider the two cases where $\alpha \leq \frac{1}{2}$ and where $\alpha > \frac{1}{2}$, and show that in each case there is a Cesaro limit distribution.

The case of $\alpha \leq \frac{1}{2}$. Let the stationary distribution of the underlying Markov chain M be $\langle \psi_1, \ldots, \psi_n \rangle$. By our assumptions about M, this distribution is independent of the start state i. When $\alpha \leq \frac{1}{2}$, we show that the $(M, \vec{\alpha}, i)$-backoff process converges to $\langle \psi_1, \ldots, \psi_n \rangle$.

Theorem 3.3 *For all states j of the $(M, \vec{\alpha}, i)$-backoff process, we have* $\lim_{t \to \infty} \Pr[X_t = j] = \psi_j$. *Thus, the limiting probability is independent of the start state i.*

Proof. Fix $\epsilon > 0$, and choose t_0 large enough that for all states j of M and all $t \geq t_0$, we have $|q_t(j) - \psi_j| < \epsilon/2$. Since $\alpha \leq 1/2$, we can also choose $t_1 \geq t_0$ large enough that for each $t \geq t_1$, we have $\Pr[Y_t > t_0] > 1 - \epsilon/2$. Then for $t \geq t_1$ we have

$$\begin{aligned}
|\Pr[X_t = j] - \psi_j| &= \left| \sum_r q_r(j) \cdot \Pr[Y_t = r] - \psi_j \sum_r \Pr[Y_t = r] \right| \\
&\leq \sum_r |q_r(j) - \psi_j| \cdot \Pr[Y_t = r] \\
&= \sum_{r < t_1} |q_r(j) - \psi_j| \cdot \Pr[Y_t = r] + \sum_{r \geq t_1} |q_r(j) - \psi_j| \cdot \Pr[Y_t = r] \\
&\leq \sum_{r < t_1} \Pr[Y_t = r] + \sum_{r \geq t_1} \epsilon/2 \cdot \Pr[Y_t = r] \\
&\leq \epsilon/2 + \epsilon/2 = \epsilon.
\end{aligned}$$

■

Although the proof above applies to each $\alpha \leq \frac{1}{2}$, we note a qualitative difference between the case of $\alpha < \frac{1}{2}$ and the "threshold case" $\alpha = \frac{1}{2}$. In the former case, for every r, there are almost surely only finitely many t for which $Y_t \leq r$; the largest such t is a step on which the process pushes a state that is never popped in the future. In the latter case, Y_t almost surely returns to 0 infinitely often, and yet the process still converges to the stationary distribution of M.

The case of $\alpha > \frac{1}{2}$. When $\alpha > \frac{1}{2}$, the $(M, \vec{\alpha}, i)$-backoff process retains positive probability on short histories as t increases, and hence retains memory of its start state i. Nevertheless, the process has a Cesaro limit distribution; but this distribution may be different from the stationary distribution of M.

Theorem 3.4 *When $\alpha > \frac{1}{2}$, the $(M, \vec{\alpha}, i)$-backoff process has a Cesaro limit distribution.*

Proof. For all natural numbers t and states j we have $\Pr[X_t = j] = \sum_r q_r(j) \cdot \Pr[Y_t = r]$ by Lemma 3.2. Viewing Y_t as a random walk on the natural numbers, one can compute the Cesaro limit of $\Pr[Y_t = r]$ to be $\zeta_r = \beta\alpha$ when $r = 0$, and $\zeta_r = \beta z^{r-1}$ when $r > 0$, where $\beta = (2\alpha - 1)/(2\alpha^2)$ and $z = (1 - \alpha)/\alpha$. (Note that Y_t does not have a stationary distribution, because it is even only on even steps.) A standard argument then shows that $\Pr[X_t = j]$ has the Cesaro limit $\sum_r \zeta_r q_r(j)$.
■

Note that the proof shows only a Cesaro limit distribution, rather than a stationary distribution. In fact, it is not hard to show that if $\alpha > \frac{1}{2}$, then there is not necessarily a stationary distribution.

Now, more generally, suppose that the process starts from an initial distribution over states; we are given a vector z_0, choose a state j with probability $z_0(j)$, and begin the process from j. As z_0 ranges over all possible probability vectors, what are the possible vectors of limiting distributions? Let us again assume a fixed underlying Markov chain M, and denote this set of limiting distributions by S_α.

Theorem 3.5 *Each S_α is a simplex. As α converges to $\frac{1}{2}$ from above, these simplices converge to the single vector that is the stationary distribution of the underlying Markov chain.*

4 Varying α's

Recall that the state space $\mathcal{S}$ of the $(M, \vec{\alpha}, i)$-Markov chain contains all finite attainable histories of the backoff process. Let us refer to the transition probability matrix of the $(M, \vec{\alpha}, i)$-Markov chain as the *Polish matrix with starting state* i, or simply the *Polish matrix* if i is implicit or irrelevant. Note that even though the backoff process has only finitely many states, the Polish matrix has a countably infinite number of states.

Our analysis in the rest of the paper will branch, depending on whether the Polish matrix is transient, null, or ergodic. We now define these concepts, which are standard notions in the study of denumerable Markov chains (see e.g., [6]). A Markov chain (and its matrix P) are called *transient* if, started in some state i, the probability of eventually returning to state i is strictly less than 1. For every irreducible[3] non-transient Markov chain P, the sequence of powers of P has a

[3]Note that the assumption that the $(M, \vec{\alpha})$-backoff process is irreducible implies that the Polish matrix is irreducible *except* if some $\alpha_i = 0$. We will see later that whenever some $\alpha_i = 0$, then the Polish matrix is transient. So all recurrent chains we encounter are irreducible.

Cesaro limit L (that is, $\frac{1}{t}\sum_{\tau=1}^{t} P^{\tau}$ converges to L). An irreducible non-transient chain is *null* if L is identically 0, and otherwise is *ergodic*. For an ergodic chain, every entry of L is strictly positive. For each state i of an ergodic chain, the expected time, starting in state i, to return to i is finite. For each state i of a null chain, the expected time, starting in state i, to return to i is infinite. We note that no finite Markov chain is null.

For example, consider a random walk on the semi-infinite line, with a reflecting barrier at 0, where the probability of moving left (except at 0) is p, of moving right (except at 0) is $1-p$, and of moving right at 0 is 1. If $p < 1/2$, then the walk is transient; if $p = 1/2$, then the walk is null; and if $p > 1/2$, then the walk is ergodic.

We say that the backoff process $(M, \vec{\alpha}, i)$ is transient (resp., null, ergodic) if the Polish matrix is transient (resp., null, ergodic). In the constant α case (Section 3), if $\alpha < 1/2$, then the backoff process is transient; if $\alpha = 1/2$, then the backoff process is null; and if $\alpha > 1/2$, then the backoff process is recurrent. The next proposition says that the classification does not depend on the start state and therefore we may refer to the backoff process $(M, \vec{\alpha})$ as being transient, ergodic, or null. Its proof may be found in Appendix C.1.

Proposition 4.1 *The backoff process $(M, \vec{\alpha}, i)$ is transient (resp., ergodic, null) precisely if the backoff process $(M, \vec{\alpha}, j)$ is transient (resp., ergodic, null).*

Theorem 4.2 *If $(M, \vec{\alpha})$ is irreducible then the task of classifying the $(M, \vec{\alpha})$-backoff process as transient or ergodic or null is solvable in polynomial time.*

Theorem 4.3 *For every irreducible $(M, \vec{\alpha})$ and for every $i \in V$, the $(M, \vec{\alpha}, i)$-backoff process has a Cesaro limit distribution. This limiting distribution is independent of i if the $(M, \vec{\alpha})$-backoff process is transient or null. Furthermore, the limiting distribution is computable exactly in polynomial time if the process is ergodic or null.*

When the $(M, \vec{\alpha})$-backoff process is transient, the limiting probabilities are not necessarily rational in the entries of M and $\vec{\alpha}$ and therefore we cannot hope to compute them exactly. In Section 4.2, we give an algorithm for approximating these limiting probabilities.

4.1 Classifying the backoff process

In this section we show how it is possible to classify, in polynomial time, the behavior of any $(M, \vec{\alpha})$-backoff process as transient or ergodic or null. In Section 3 (where the backoff probability is independent of the state), we showed that the length of the history is either always expected to grow or always expected to shrink (except for initial histories), independent of the top state in the history stack. To see that this argument cannot carry over to this section, consider a simple Markov chain M on two states with $M_{ij} = 1/2$ for every pair i, j and $\vec{\alpha}_1 = \langle .99, .01 \rangle$. It is clear that if the top state is 1, then the history is expected to shrink while if the top state is 2, then the history is expected to grow. To deal with this imbalance between the states, we associate a weight w_i with every state i and consider the weighted sum of states on the stack. Our goal is to find a weight vector with the property that the weighted sum of states on the stack is expected to grow (resp. shrink) if and only if the history is expected to grow unboundedly (resp. remain bounded). This hope motivates our next few definitions.

Definition 4.4 *For a nonnegative vector $\vec{w} = \langle w_1, \ldots, w_n \rangle$, and a history $\bar{\sigma} = \langle \sigma_0, \ldots, \sigma_l \rangle$ of an $(M, \vec{\alpha})$-backoff process on n states define the w-potential of $\bar{\sigma}$, denoted $\Phi_{\vec{w}}(\bar{\sigma})$, to be $\sum_{i=1}^{l} w_{\sigma_i}$ (i.e., it is the weighted sum of the states in the history, except the initial state, with state i weighted by w_i).*

Definition 4.5 *For a nonnegative vector $\vec{w}$, and a history $\bar{\sigma}$ of an $(M, \vec{\alpha})$-backoff process on n states define the $\vec{w}$-differential of $\bar{\sigma}$, denoted $\Delta\Phi_{\vec{w}}(\bar{\sigma})$, to be $\mathrm{E}\left[\Phi_{\vec{w}}(\mathrm{succ}(\bar{\sigma}))\right] - \Phi_{\vec{w}}(\bar{\sigma})$. (Here E represents the expected value over the distribution given by $\mathrm{succ}(\bar{\sigma})$.)*

The following proposition is immediate from the definition.

Proposition 4.6 *If $\bar{\sigma}$ and $\bar{\sigma}'$ are non-initial histories with the same top state j, then*

$$\Delta\Phi_{\vec{w}}(\bar{\sigma}) = \Delta\Phi_{\vec{w}}(\bar{\sigma}') = -\alpha_j w_j + (1 - \alpha_j) \sum_{k=1}^{n} M_{jk} w_k.$$

The above proposition motivates the following definition.

Definition 4.7 *For an $(M, \vec{\alpha})$-backoff process, nonnegative vector $\vec{w}$, and state $j \in \{1, \ldots, n\}$, let $\Delta\Phi_{\vec{w},j} = \Delta\Phi_{\vec{w}}(\bar{\sigma})$, where $\bar{\sigma}$ is any history with $j = \mathrm{top}(\bar{\sigma})$ and $\ell(\bar{\sigma}) > 0$. Let $\Delta\Phi_{\vec{w}}$ denote the vector $\langle \Delta\Phi_{\vec{w},1}, \ldots, \Delta\Phi_{\vec{w},n} \rangle$.*

For intuition, consider the constant α case with weight vector $w_i = 1$ for all i. In this case $\Phi_{\vec{w}}(\bar{\sigma})$, the w-potential of $\bar{\sigma}$, is precisely $\ell(\bar{\sigma})$, and $\Delta\Phi_{\vec{w}}(\bar{\sigma})$, the $\vec{w}$-differential of $\bar{\sigma}$, is the expected change in the size of the stack, which is $1 - 2\alpha$. When $\alpha < 1/2$ (resp., $\alpha = 1/2$, $\alpha > 1/2$), so that the expected change in the size of the stack is positive (resp., 0, negative), the process is transient (resp., null, ergodic).

Similarly, in the non-constant α case we would like to associate a positive weight with every state so that (1) the expected change in potential in every step has the same sign independent of the top state (i.e., $\vec{w}$ is positive and $\Delta\Phi_{\vec{w}}$ is either positive or zero or negative), and (2) this sign can be used to categorize the process as either transient, null or ergodic precisely as it did in the constant α case.

In general, this will not be possible, say, if some $\alpha_i = 1$ and some other $\alpha_j = 0$. Therefore, we relax this requirement slightly and define the notion of an "admissible" vector (applicable to both the vector of weights and also the vector of changes in potential).

Definition 4.8 *We say that an n-dimensional vector $\vec{v}$ is admissible for a vector $\vec{\alpha}$ if $\vec{v}$ is non-negative and $v_i = 0$ only if $\alpha_i = 1$. (We will drop the suffix, admissible for $\vec{\alpha}$, and simply say admissible, if the latter vector is named $\vec{\alpha}$.)*

In Appendix C.3 we prove three very natural lemmas that combine to show the following. Given $(M, \vec{\alpha})$ and an admissible vector $\vec{w}$: (1) If $\Delta\Phi_w$ is admissible then the process is transient. (2) If $\Delta\Phi_w$ is zero then the process is null. (3) If $-\Delta\Phi_w$ is admissible then the process is ergodic. Roughly speaking, we show that $\Phi_{\vec{w}}(\bar{\sigma})$ is a bounded-difference martingale. This enables us to use martingale tail inequalities to analyze the long-term behavior of the process.

This explains what could happen if we are lucky with the choice of $\vec{w}$. It does not explain how to find $\vec{w}$, or even why the three cases above are exhaustive. Our next lemma shows that the cases are indeed exhaustive and gives a efficient algorithm to compute $\vec{w}$.

Lemma 4.9 *For every irreducible $(M, \vec{\alpha})$-backoff process, there exists an admissible vector $\vec{w}$ such that exactly one of the following holds: (1) $\Delta\Phi_{\vec{w}}$ is admissible, (2) $\Delta\Phi_{\vec{w}}$ is zero, or (3) $-\Delta\Phi_{\vec{w}}$ is admissible. Furthermore such a vector can be computed in polynomial time, given $(M, \vec{\alpha})$.*

Proof. We first get rid of an easy case, namely if some $\alpha_j = 0$.

Claim 4.10 *Suppose there exists j such that $\alpha_j = 0$. Let $d(k,j)$ denote the length of a shortest path (of non-zero probability) from k to j in M. Let $M_{\min}$ be the smallest non-zero entry of M, and let $\alpha_{\max}$ be the largest entry of $\vec{\alpha}$ that is strictly smaller than 1. Let $\gamma = \frac{1}{2} \cdot M_{\min} \cdot \frac{1-\alpha_{\max}}{\alpha_{\max}}$. Let $\vec{w}$ be defined as follows.*

$$w_k = \begin{cases} 0 & \text{if } \alpha_k = 1 \\ \gamma^{d(k,j)} & \text{otherwise.} \end{cases}$$

Then $\vec{w}$ and $\Delta\Phi_{\vec{w}}$ are admissible.

We defer the proof of the claim to Appendix C.2.

Let A be the $n \times n$ diagonal matrix with the $A_{ii} = \alpha_i$. Let I be the $n \times n$ identity matrix. Then notice that $\Delta\Phi_{\vec{w}} = -A\vec{w} + (I - A)M\vec{w}$.

Let $H = (I - A)MA^{-1}$. Notice that since none of the α_i's are zero, A^{-1} exists. The matrix H is nonnegative. Let $H|_{\vec{\alpha}}$ be the restriction of H to rows and columns corresponding to $\alpha_j < 1$. Notice that $H|_{\vec{\alpha}}$ is irreducible. (This is equivalent to $M|_{\vec{\alpha}}$ being irreducible, which is implied by the irreducibility of the backoff process.) By the Perron-Frobenius theorem (Theorem A.1), there exists a (unique) positive vector $\vec{v}'$ and a (unique) positive real ρ such that $H|_{\vec{\alpha}}\vec{v}' = \rho\vec{v}'$. Let $\vec{v}$ be an n-dimensional vector obtained by padding $\vec{v}'$ with zeroes in the columns corresponding to $\alpha_j = 1$. Notice then that $H\vec{v} = \rho\vec{v}$ and $\vec{v}$ is admissible for $\vec{\alpha}$. Now let $\vec{w} = A^{-1}\vec{v}$. Notice that $\vec{w}$ is also admissible, and satisfies $(I - A)M\vec{w} = \rho A\vec{w}$. Equivalently, $-A\vec{w} + (I - A)M\vec{w} = (\rho - 1)A\vec{w}$, and thus $\Delta\Phi_{\vec{w}} = (\rho - 1)A\vec{w}$. Thus, for this choice of $\vec{w}$, (1) if $\rho > 1$, then $\Delta\Phi_{\vec{w}}$ is admissible; (2) if $\rho = 1$, then $\Delta\Phi_{\vec{w}} = 0$; (3) if $\rho < 1$, then $-\Delta\Phi_{\vec{w}}$ is admissible.

Thus we have proved the existence part of the result. But it also follows that the vector $\vec{w}$ can be computed efficiently (since this amounts to computing an eigenvalue and the corresponding eigenvector of a given matrix). ∎

It is easy to see that the results we have discussed combine to prove Theorem 4.2.

4.2 Cesaro limit distributions

We begin the section by sketching the proof that the $(M, \vec{\alpha}, i)$-backoff process always has a Cesaro limit distribution. The proof is quite different in each of the cases (transient, ergodic and null). More details appear in Appendix C.4. We conclude the section by showing how the limiting distribution may be computed.

The easiest case is the ergodic case. Since the Polish matrix is ergodic, the corresponding Markov process has a Cesaro limit. This gives us a Cesaro limit in the backoff process, where the probability of state i is the sum of the probabilities of the stacks in the Polish matrix with top state i.

We now consider the transient case. When the backoff process is in a state (with a given stack), and that state is never popped off of the stack (by taking a backedge), then we refer to this (occurrence of the) state as *irrevocable*. Let us fix a state i, and consider a renewal process (see Definition A.7), where each new epoch begins when the process has an irrevocable occurrence of state i. Since the Polish matrix is transient, the expected length of an epoch is finite. The limiting probability distribution of state j is the expected number of times that the process is in state j in an epoch, divided by the expected length of an epoch. This is a sketch of a proof of the existence of a Cesaro limit distribution. A more careful argument (given in Appendix C.4) shows the existence of a stationary distribution.

Finally, we consider the null case. We select a state j where $\alpha_j \neq 1$. Let us consider a new backoff process, where the underlying Markov matrix M is the same; where all of the backoff probabilities α_k are the same, except that we change α_j to 1; and where we change the start state to j. This new backoff process can be shown to be ergodic. We show a way of "pasting together" runs of the new ergodic backoff process to simulate runs of the old null ergodic process. Thereby, we show the remarkable fact that the old null process has a Cesaro limit distribution which is the same as the Cesaro limit distribution of the new ergodic process. We now show how the limiting distribution may be computed. Again, we branch into three cases.

4.2.1 The null case

The matrix $H = (I - A)MA^{-1}$, which we saw in Section 4.1), plays an important role in this section. We refer to this matrix as the *Hungarian matrix* of the $(M, \vec{\alpha})$-backoff process. The next theorem gives an important application of the Hungarian matrix.

Theorem 4.11 *The limiting probability distribution π satisfies $\pi = \pi H$. This linear system has a unique solution subject to the restriction $\sum_i \pi_i = 1$. Thus, the limiting probability distribution can found by solving a linear system.*

Proof [Sketch]. The key ingredient in the proof is the observation that in the null case, the probability of a transition from a state i to a state j by a forward step is the same as the probability of a transition from state j to a state i by a backward step (since each forward move is eventually revoked, with probability 1). Thus if we let $\pi_{i \to j}$ denote the probability of a forward step from i to j and $\pi_{i \leftarrow j}$ denote the probability of a backward step from j to i (and π_i denotes the limiting probability of being in state i), then the following conditions hold:

$$\pi_i = \sum_j \pi_{i \to j} + \sum_j \pi_{j \leftarrow i}; \qquad \pi_{i \to j} = (1 - \alpha_i) M_{ij} \pi_i; \qquad \pi_{i \to j} = \pi_{i \leftarrow j}.$$

Manipulating the above shows that π satisfies $\pi = \pi H$. For the uniqueness part, notice that if all $\alpha_i < 1$, then H is irreducible and nonnegative and thus by Theorem A.1, π is a maximal eigenvector and hence a unique solution to the linear system. If some $\alpha_i = 1$, we argue by focusing on the matrix $H|_\alpha$, (as in Section 4.1, $H|_\alpha$ is the principal submatrix of H containing only rows and columns corresponding to i s.t. $\alpha_i < 1$) which is irreducible. Details omitted. ■

4.2.2 The ergodic case

In this case also the limiting probabilities are obtained by solving linear systems, obtained from a renewal argument. We define "epochs" starting at i by simulating the backoff process as follows. The epoch starts at an initial history with $X_0 = \langle i \rangle$. At the first step the process makes a forward step. At every subsequent unit of time, if the process is in state j, it first flips a coin that comes up "B" with probability α_j and "F" otherwise. If the coin comes up "B", it checks to see if it is back at the initial history and if so declares an end of an epoch.

Notice that the distribution of the length of an epoch starting at i is precisely the same as the distribution of time, starting at an arbitrary history with i on top of the stack, until this occurrence of i is popped from the stack, conditioned on the fact that the first step taken from i is a forward step.

Let T_i denote the expected length of (or more precisely, number of transitions in) an epoch, when starting at state i. Let N_{ij} denote the expected number of transitions out of state j in an epoch when starting at state i. Standard renewal arguments (using Theorem A.8 with $E(X_i) = T_i$

and $E(Y_i) = N_{ij}$) show that the Cesaro limit probability distribution vector $\pi^{(i)}$, for an $(M, \vec{\alpha}, i)$-backoff process, is given by $\pi_j^{(i)} = N_{ij}/T_i$, provided T_i is finite. Since this is true for the ergodic case, this gives us a way to compute the Cesaro limit distribution in the ergodic case. The key equations that allow us to compute the N_{ij} and T_i are:

$$T_i = 1 + \sum_k M_{ik}[\alpha_k \cdot 1 + (1-\alpha_k)(T_k + 1)] + (1-\alpha_i)T_i,$$

$$N_{ij} = \delta_{ij} + \sum_k M_{ik}[\alpha_k \cdot \delta_{jk} + (1-\alpha_k)(N_{kj} + \delta_{jk})] + (1-\alpha_i)N_{ij},$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. (The above equations are derived by straightforward conditioning. For example, if the first step in the epoch takes the process to state k, then it takes T_k units of time to return to $\langle i \rangle$ and then with probability $(1-\alpha_i)$ it takes T_i more steps to end the epoch.)

The T_i's and N_{ij}'s can be computed by solving the linear systems above. Uniqueness of the solution to the linear system is derived from the non-singularity of $I - H$, where $H = (I - A)MA^{-1}$ is the Hungarian matrix. Details omitted.

4.2.3 The transient case

Theorem 4.12 *Let $(M, \vec{\alpha})$ be a transient backoff process on n states, and let all entries of M and $\vec{\alpha}$ be rationals expressible as ratios of l-bit integers. Then, given any error bound $\epsilon > 0$, a vector π' that ϵ-approximates the stationary probability distribution π of the $(M, \vec{\alpha})$-process (i.e., satisfies $|\pi'_j - \pi_j| \leq \epsilon$) can be computed in time polynomial in n, l and $\log \frac{1}{\epsilon}$.*

Let r_i denote the "revocation" probability of a state i, i.e., the probability that an epoch starting at i, as in Section 4.2.2, ends in finite time. Let $\vec{r}$ denote the vector of revocation probabilities. The following lemma shows how to compute the limiting probabilities π given $\vec{r}$. Further it shows how to compute a close approximation to π, given a sufficiently close approximation to $\vec{r}$.

Lemma 4.13 *The limiting probabilities satisfy $\pi = \pi(I - A)MR$, where R is a diagonal matrix with $\frac{1}{1-(1-\alpha_i)\sum_k r_k M_{ik}}$ as the ith entry. Further, there exists a unique solution to the above system subject to the condition $\sum_i \pi_i = 1$.*

Remark: Two special cases of interest are: (1) When all $\alpha_i = 0$, in which case we recover the familiar condition for Markov chains that $\pi = \pi M$. (2) When all $r_i = 1$, in which case we get the null case and then π satisfies $\pi = \pi(I - A)MA^{-1}$.

Proof. The first part of the lemma is obtained as in Theorem 4.11. Let $\pi_{i \to j}$ denote the probability of a forward transition from i to j at stationarity, and let $\pi_{i \leftarrow j}$ denote the probability of a backward transition from j to i. Then the following conditions hold.

$$\pi_{i \leftarrow j} = r_j \pi_{i \to j} \tag{1}$$

$$\pi_{i \to j} = \pi_i (1-\alpha_i) M_{ij} \tag{2}$$

$$\pi_i = \sum_j \pi_{j \to i} + \sum_j \pi_{i \leftarrow j} \tag{3}$$

Using Equation (1) to eliminate all occurrences of variables of the form $\pi_{i \leftarrow j}$, and then Equation (2) to eliminate all occurrences of $\pi_{i \to j}$, Equation (3) simplifies to:

$$\pi_i = \sum_j \pi_j (1-\alpha_j) M_{ji} + \pi_i (1-\alpha_i) \sum_j r_j M_{ij} \tag{4}$$

Thus if we let D be the matrix with the ijth entry being

$$D_{ij} = \frac{(1-\alpha_i)M_{ij}}{1-(1-\alpha_j)\sum_k M_{jk}r_k}$$

then π satisfies $\pi = \pi D$. As in the proof of Theorem 4.11 if we permute the rows and columns of D so that all states i with $\alpha_i = 1$ appear at the end, then the matrix D looks as follows:

$$D = \begin{bmatrix} D_\alpha & X \\ 0 & 0 \end{bmatrix}$$

where D_α is non-negative and irreducible. Thus $\pi = [\pi_A \pi_B]$ must satisfy $\pi_A = \pi_A D_\alpha$ and $\pi_B = \pi_A X$. π_A is now seen to be unique (up to scaling) by the Perron-Frobenius theorem (Theorem A.1), while π_B is unique given π_A. The lemma follows by noticing that D can be expressed as $(I-A)MR$. ■

Lemma 4.14 *Let the entries of M and $\vec{\alpha}$ be l-bit rationals describing a transient $(M,\vec{\alpha})$-precess and let π be its stationary probability vector. For every $\epsilon > 0$, there exists $\beta > 0$, with $\log\frac{1}{\beta} = \mathrm{poly}(n, l, \log\frac{1}{\epsilon})$, such that given any vector $\vec{r'}$ of l'-bit rationals satisfying $\|\vec{r'} - \vec{r}\|_\infty \le \beta$, a vector π' satisfying $\|\pi' - \pi\|_\infty \le \epsilon$ can be found in time $\mathrm{poly}(n, l, l', \log\frac{1}{\epsilon})$.*

Remark: By truncating $\vec{r'}$ to $\log\frac{1}{\beta}$ bits, we can ensure that l' also grows polynomially in the input size, and thus get a fully polynomial time algorithm to approximate π.

We defer the proof of Lemma 4.14 to Appendix C.5.

Lemma 4.15 *The revocation probabilities r_i are the optimum solution to the following system:*

$$\left.\begin{array}{lrcl} & \min \sum_i x_i & & \\ s.t. & x_i & \ge & \alpha_i + (1-\alpha_i)x_i \sum_j M_{ij}x_j \\ & x_i & \le & 1 \\ & x_i & \ge & 0 \end{array}\right\} \quad (5)$$

Further, the system of inequalitities above is equivalent to the following semidefinite program:

$$\left.\begin{array}{lrcl} & \min \sum_i x_i & & \\ s.t. & q_i & = & 1-(1-\alpha_i)\sum_j M_{ij}x_j \\ & x_i & \le & 1 \\ & x_i & \ge & 0 \\ & q_i & \ge & 0 \\ & D_i & & \textit{postive semidefinite, where } D_i = \begin{bmatrix} r_i & \sqrt{\alpha_i} \\ \sqrt{\alpha_i} & q_i \end{bmatrix} \end{array}\right\} \quad (6)$$

Proof. We start by considering the following iterative system and prove that it converges to the optimum of (5).

For $t = 0, 1, 2, \ldots$, define $x_i^{(t)}$ as follows:

$$x_i^{(0)} = 0, \quad x_i^{(t+1)} = \alpha_i + (1-\alpha_i)x_i^{(t)} \sum_j M_{ij}x_j^{(t)}.$$

By induction, we note that $x_i^{(t)} \leq x_i^{(t+1)} \leq 1$. For example:

$$\begin{aligned} x_i^{(t+1)} &= \alpha_i + (1-\alpha_i)x_i^{(t)} \sum_j M_{ij} x_j^{(t)} \\ &\geq \alpha_i + (1-\alpha_i)x_i^{(t-1)} \sum_j M_{ij} x_j^{(t-1)} \\ &= x_i^{(t)} \end{aligned}$$

Thus since $\langle x_i^{(t)} \rangle_t$ is a non-decreasing sequence in the interval $[0,1]$, it must have a limit. Let x_i^* denote this limit.

We claim that x_i^* are the (unique) optimum to (5). By construction, it is clear that $0 \leq x_i^* \leq 1$ and $x_i^* = \alpha_i + (1-\alpha_i)x_i^* \sum_j M_{ij} x_j^*$; and hence x_i^*'s form a feasible solution to (5). To prove it is the optimum, we claim for every feasible solution a_i's to (5) satisfies $a_i \geq x_i^{(t)}$ and thus $a_i \geq x_i^*$. We prove this claim by induction. Assume $a_i \geq x_i^{(t)}$, for every i. Then

$$\begin{aligned} a_i &\geq \alpha_i + (1-\alpha_i)a_i \sum_j M_{ij} a_j \\ &\geq \alpha_i + (1-\alpha_i)x_i^{(t)} \sum_j M_{ij} x_j^{(t)} \\ &= x_i^{(t+1)}. \end{aligned}$$

This concludes the proof that the x_i^*'s are the unique optimum to (5).

Next we show that the revocation probability $r_i = x_i^*$. To do so, note first that r_i satisfies the condition

$$r_i = \alpha_i + (1-\alpha_i) \sum_j M_{ij} r_j r_i.$$

(Either the move onto i is revoked at the first step with probability α_i, or we move to j with probability $(1-\alpha_i)M_{ij}$ and then the move to j is eventually revoked with probability r_j, and this places i again at the top of the stack, and with probability r_i this move is revoked eventually.) Thus r_i's form a feasible solution and thus $r_i \geq x_i^*$. To prove that $r_i \leq x_i^*$, let us define $r_i^{(t)}$ to be the probability that a forward move onto vertex i is revoked in at most t steps. Note that $r_i = \lim_{t\to\infty} r_i^{(t)}$. We will show by induction that $r_i^{(t)} \leq x_i^{(t)}$ and this implies $r_i \leq x_i^*$. Notice first that

$$r_i^{(t+1)} \leq \alpha_i + (1-\alpha_i) \sum_j M_{ij} r_j^{(t)} r_i^{(t)}.$$

(This follows from a conditioning argument similar to the above and then noticing that in order to revoke the move within $t+1$ steps, both the revocation of the move to j and then the eventual revocation of the move to i must occur within t time steps.) Now an inductive argument as earlier shows $r_i^{(t+1)} \leq x_i^{(t+1)}$. Thus we conclude that $x_i^* = r_i$. This finishes the first part of the lemma.

For the second part, note that condition D_i is semidefinite is equivalent to the condition $r_i q_i \geq \alpha_i$. Substituting $q_i = 1-(1-\alpha_i)\sum_j M_{ij} r_j$ turns this into the constraint $r_i - (1-\alpha_i) r_i \sum_j M_{ij} r_j \geq \alpha_i$, and thus establishing the (syntactic) equivalence of (5) and (6). ■

Lemma 4.16 *If the entries of M and $\vec{\alpha}$ are given by l-bit rationals, then an ϵ-approximation to the vector of revocation probabilities can be found in time* $\mathrm{poly}(n, l, \log \frac{1}{\epsilon})$.

Proof. We solve the convex program given by (5) approximately using the ellipsoid algorithm [3]. Recall that the ellipsoid algorithm can solve a convex programming problem given (1) a separation oracle describing the convex space, (2) a point $\vec{x}$ inside the convex space, (3) radii ϵ and R such that the ball of radius ϵ around $\vec{x}$ is contained in the convex body and the ball of radius R contains the convex body. The running time is polynomial in the dimension of the space and in $\log \frac{R}{r}$.

The fact that (5) describes a convex program follows from the fact that it is equivalent to the semidefinite program (6). Further, a separation oracle can also be obtained due to this equivalence. In what follows we will describe a vector $\vec{x}$ that is feasible, and an $\epsilon \geq 2^{-\operatorname{poly}(n,l)}$ such that every point y satisfying $\|x - y\|_\infty \leq \epsilon$ is feasible. Further it is trivial to see that every feasible point satisfies the condition that the ball of radius $\sqrt{n}$ around it contains the unit cube and hence all feasible solutions. This will thus suffice to prove the claim.

Recall, from Section 4.1, that since $(M, \vec{\alpha})$ is null, there exists a $\rho > 1$ and a vector $\vec{w}$ such that $(I - A)MA^{-1}(\vec{w}) \geq \rho\vec{w}$. (This held even in the case where some $\alpha_j = 0$.) Note further that since ρ is the maximal eigenvalue of a matrix whose entries are $\operatorname{poly}(n, l)$ bit rationals, its value is at least $1 + 2^{-\operatorname{poly}(n,l)}$. Let $\vec{v} = A^{-1}\vec{w}$, and let $v_{\max} = \max_i v_i$ and let $v_{\min} = \min_i v_i$. Again we note that $v_{\min} \geq 2^{-\operatorname{poly}(n,l)} v_{\max}$. Scaling $\vec{v}$ appropriately, we may assume $v_{\max} = 1$. We will use this ρ and $\vec{v}$ below.

Before describing the vector $\vec{x}$ and ϵ, we make one simplification. Notice that if $\alpha_i = 1$, then $r_i = 1$ and if $\alpha_i = 0$, then $r_i = 0$. We fix this setting and then solve (5) for only the remaining choices of indices i. So henceforth we assume $0 < \alpha_i < 1$ and in particular the fact that $\alpha_i \geq 2^{-l}$.

Let $\delta = \frac{\rho-1}{2\rho}$. Note $\delta > 2^{-\operatorname{poly}(n,l)}$. Let $\epsilon = 2^{-(l+3)} v_{\min} \left(\frac{\rho-1}{\rho}\right)^2$. We will set $z_i = 1 - \delta v_i$ and show that $z_i - \alpha_i - (1 - \alpha_i) z_i \sum_j M_{ij} z_j$ is at least 2ϵ. Now letting $x_i = z_i - \epsilon$, we get the required vector $\vec{x}$ and ϵ.

Consider

$$\begin{aligned}
& z_i - \alpha_i - (1 - \alpha_i) z_i \sum_j M_{ij} z_j \\
&= 1 - \delta v_i - \alpha_i - (1 - \alpha_i)(1 - \delta v_i) \sum_j M_{ij}(1 - \delta v_j) \\
&= 1 - \delta v_i - \alpha_i - (1 - \alpha_i)(1 - \delta v_i)(1 - \delta \sum_j M_{ij} v_j) \\
&= (1 - \delta v_i) \left(\delta \sum_j (1 - \alpha_i) M_{ij} v_j \right) - \delta v_i \alpha_i \\
&\geq (1 - \delta v_i)(\delta \rho \alpha_i v_i) - \delta v_i \alpha_i \\
&\geq \delta \alpha_i v_i (\rho - \rho \delta v_i - 1) \\
&\geq \left(\frac{\rho - 1}{2\rho} \right)^2 \alpha_i v_i \\
&\geq 2\epsilon.
\end{aligned}$$

This concludes the proof. ■

Proof. **[of Theorem 4.12]** Given M, $\vec{\alpha}$ and ϵ, let β be as given by Lemma 4.14. We first compute a β approximation to the vector of revocation probabilities in time $\operatorname{poly}(n, l, \log \frac{1}{\beta}) = \operatorname{poly}(n, l, \log \frac{1}{\epsilon})$ using Lemma 4.16. The output is a vector $\vec{r'}$ of $l' = \operatorname{poly}(n, l, \log \frac{1}{\epsilon})$-bit rationals. Applying Lemma 4.14 to M, $\vec{\alpha}$, $\vec{r}$ and ϵ, we obtain a ϵ-approximation to the stationary probability vector π in time $\operatorname{poly}(n, l, l', \log \frac{1}{\epsilon}) = \operatorname{poly}(n, l, \log \frac{1}{\epsilon})$. ■

References

[1] R. Barrett, P.P. Maglio, and D.C. Kellem. How to personalize the web. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '97).*

[2] M. Charikar, S.R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins. On targeting Markov segments. *Proceedings of the ACM Symposium on Theory of Computing, 1999.*

[3] M. Grotschel, L. Lovasz, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimizations.* Springer-Verlag, Berlin, 1987.

[4] R.A. Horn and C.A. Johnson. *Matrix Analysis*, Cambridge University Press, 1985.

[5] S. Karlin and H. Taylor *A First Course in Stochastic Processes*, Academic Press, 1975.

[6] J.G. Kemeny, J. L. Snell, and A.W. Knapp. *Denumerable Markov Chains.* The University Series in Higher Mathematics. Van Nostrand, Princeton, NJ, 1966.

[7] H. Lieberman. An agent that assists web browsing. *Proceedings of IJCAI*, 1995.

[8] R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, Cambridge, 1994.

[9] T. Selker. COACH: a teaching agent that learns. *Communications of the ACM* 37:7, 1994.

A Preliminaries

In this section, we review background material essential to our proofs.

A.1 Perron-Frobenius Theorem

Theorem A.1 *(Perron-Frobenius Theorem, see e.g., [4], p. 508) Let A be an irreducible, non-negative n by n matrix. Then*

- *There exists $\vec{v}$, with all components positive, and $\lambda_0 > 0$ such that $A\vec{v} = \lambda_0\vec{v}$;*
- *if $\lambda \neq \lambda_0$ is any other eigenvalue of A, then $|\lambda| < \lambda_0$; and*
- *the eigenspace associated with λ_0 is one-dimensional.*

A.2 Martingale Tail Inequalities

We begin by reviewing the basic definitions.

Definition A.2

- *A sequence of random variables $X_0, X_1, \ldots$ is said to be a* martingale *if for all $i > 0$, $E[X_i|X_0, \ldots, X_{i-1}] = X_{i-1}$.*
- *A sequence of random variables $X_0, X_1, \ldots$ is said to be a* super-martingale *if for all $i > 0$, $E[X_i|X_0, \ldots, X_{i-1}] \leq X_{i-1}$.*
- *A sequence of random variables $X_0, X_1, \ldots$ is said to be a* sub-martingale *if for all $i > 0$, $E[X_i|X_0, \ldots, X_{i-1}] \geq X_{i-1}$.*

Theorem A.3 *(Azuma's Inequality, see e.g., [8], p. 92) Let $X_0, X_1, \ldots$ be a martingale sequence such that for each k*

$$|X_k - X_{k-1}| \leq c_k,$$

where c_k may depend on k. Then, for all $t \geq 0$ and any $\lambda > 0$,

$$\Pr[X_t - X_0| \geq \lambda] \leq 2exp\left(-\frac{\lambda^2}{2\sum_{1\leq k\leq t} c_k^2}\right).$$

Corollary A.4 *Let $X_0, X_1, \ldots$ be a martingale sequence such that for all k*

$$|X_k - X_{k-1}| \leq c.$$

Then, for all $t \geq 0$ and any $\lambda > 0$

$$\Pr[X_t - X_0| \geq \lambda c\sqrt{t}] \leq 2e^{-\lambda^2/2}.$$

Corollary A.5 *Let $X_0, X_1, \ldots$ be a sub-martingale such that*

$$E(X_i|X_0, \ldots, X_{i-1}) \geq X_{i-1} + \beta,$$

$(\beta > 0)$ and for all k

$$|X_k - X_{k-1}| \leq c.$$

Then, for all $t \geq 0$ and any $\lambda \geq 0$

$$\Pr(|X_t - X_0| \leq \lambda) \leq 2e^{-\left(\frac{\beta}{2c^2}\left(t - \frac{2\lambda}{\beta}\right)\right)}.$$

Corollary A.6 *Let $X_0, X_1, \ldots$ be a super-martingale such that*

$$E(X_i|X_0, \ldots, X_{i-1}) \leq X_{i-1} - \beta,$$

$(\beta > 0)$ and for all k

$$|X_k - X_{k-1}| \leq c.$$

Then, for all $t \geq 0$

$$\Pr(|X_t + \beta t - X_0| \geq \gamma t) \leq 2e^{-\gamma^2 t/(2c^2)}.$$

A.3 Renewal Theory

Definition A.7 *A* renewal process *$\{N(t),\ t \geq 0\}$ is a non-negative integer-valued stochastic process that registers the successive occurrences of an event during the time interval (0,t], where the times between consecutive events are positive, independent, identically-distributed random variables.*

Theorem A.8 *(Corollary of Renewal Theorem, see e.g., [5], p. 203) Let $N(t)$ be a renewal process where the time between the ith and $(i+1)$st event is denoted by the random variable X_i. Let Y_i be a cost or value associated with the ith renewal cycle, where the values Y_i, $i \geq 1$, are also positive, independent, identically-distributed random variables. Then*

$$lim_{t\to\infty} \frac{E[\sum_{1\leq k\leq N(t)+1} Y_k]}{t} = \frac{E(Y_1)}{E(X_1)}.$$

A.4 A Basic Probability Fact

Proposition A.9 *Let $p : \mathcal{Z}^+ \to [0,1]$ be a probability distribution (i.e., $\sum_{i=1}^{\infty} p(i) = 1$) with expectation at least μ (i.e., $\sum_{i=1}^{\infty} ip(i) \geq \mu$). Let $Y_1, \ldots, Y_N, \ldots,$ be a sequence of independent random variables distributed according to p. Then, for every $\delta > 0$ and $\mu' < \mu$, there exists an index N such that*

$$\Pr[\sum_{i=1}^{N} Y_i > \mu' \cdot N] \geq 1 - \delta.$$

Proof. Let K be an integer such that $\sum_{i=1}^{K} ip(i) \geq \frac{\mu'+\mu}{2}$. (Such an index exists since the sequence $\mu_1, \mu_2, \ldots, \mu_j, \ldots,$ where $\mu_j = \sum_{i=1}^{j} ip(i)$ is monotone non-decreasing and achieves a value of at least μ.) Let p' be the distribution with $p'(i) = p(i)$ if $i \in \{1, \ldots, K\}$, $p'(0) = 1 - \sum_{i=1}^{K} p(i)$ and 0 everywhere else. Note that the expectation of p' is at least $\frac{\mu'+\mu}{2}$, and all moments of p' are bounded (by K). Thus we can apply the law of large numbers to conclude that there exists an N such that sampling n times from p' gives an average of at least $\mu' \cdot N$ with probability at least $1 - \delta$. ■

B Proof of Theorem 3.1

Theorem B.1 *For given natural numbers λ and t, and a state j, we have* $\Pr[X_t = j \mid \ell(H_t) = \lambda] = q_\lambda(j)$.

Proof. Consider a string ω of "F"s and "B"s with the property the in every prefix, the number of "B"s is not more than the number of "F"s. Notice that every such string corresponds to a legitimate auxiliary sequence for the backoff process (except if $\alpha_i = 0$ or 1). Now consider strings ω and ω' such that $\omega = \omega_1 F B \omega_2$ and $\omega' = \omega_1 \omega_2$. Let ω be of length t and ω_1 of length t_1. Notice that

$$\begin{aligned}
&\Pr[X_t = j \mid \langle S_1, \ldots, S_t \rangle = \omega] \\
&= \sum_{\bar\sigma \in \mathcal{S}} \Pr[H_{t_1} = \bar\sigma \mid \langle S_1, \ldots, S_{t_1} \rangle = \omega_1] \cdot \Pr[X_t = j \mid \langle S_{t_1+1}, \ldots, S_{t_1} \rangle = FB\omega_2 \text{ and } H_{t_1} = \bar\sigma] \\
&= \sum_{\bar\sigma \in \mathcal{S}} \Pr[H_{t_1} = \bar\sigma \mid \langle S_1, \ldots, S_{t_1} \rangle = \omega_1] \cdot \Pr[X_t = j \mid \langle S_{t_1+3}, \ldots, S_{t_1} \rangle = \omega_2 \text{ and } H_{t_1+2} = \bar\sigma] \\
&= \Pr\left[X_{t-2} = j \mid \langle S_1, \ldots, S_{t-2} \rangle = \omega'\right]
\end{aligned}$$

This motivates the following notion of a reduction. A sequence ω of "F"s and "B"'s reduces in one step to a sequence ω' if $\omega = \omega_1 F B \omega_2$ and $\omega' = \omega_1 \omega_2$. A sequence ω reduces to a sequence ω'' if if ω'' can be obtained from ω by a finite number of "reductions in one step". Repeatedly applying the claim from the previous paragraph, we find that if a string ω of length t reduces to a string ω'' of length t'', then

$$\Pr[X_t = j \mid \langle S_1, \ldots, S_t \rangle = \omega] = \Pr[X_{t''} = j \mid \langle S_1, \ldots, S_{t''} \rangle = \omega''].$$

But every auxiliary sequence $\langle S_1, \ldots, S_t \rangle$ can eventually be reduced to a sequence of the form F^λ (i.e., consisting only of forward steps), and further $\lambda = \ell(H_t)$. This yields:

$$\Pr[X_t = j \mid \langle S_1, \ldots, S_t \rangle, \ell(H_t) = \lambda] = \Pr\left[X_\lambda = j \mid \langle S_1, \ldots, S_\lambda \rangle = F^\lambda\right] = q_\lambda(j).$$

■

C Proofs for Section 4

C.1 Preliminaries

Proof [of Proposition 4.1]. Let us call a state i *transient* if $(M, \vec{\alpha}, i)$ is transient, and similarly for the other properties (recurrent, and its subclassifications ergodic and null). We must show that if some state is transient (resp., ergodic, null) then every state is transient (resp., ergodic, null). If $\alpha_j = 0$ for some j, then every state i is transient. This is because starting in state i, there is a positive probability of eventually reaching state j, and the stack $\langle i, \ldots, j \rangle$ can never be unwound back to the original stack $\langle i \rangle$. So assume that $\alpha_j > 0$ for every j.

Assume that there is at least one transient state and at least one recurrent state; we shall derive a contradiction. Assume first that there is some transient state j with $\alpha_j < 1$. Let i be a recurrent state. Starting in state i, there is a positive probability of eventually reaching state j. This gives the stack $\langle i, \ldots, j \rangle$. There is now a positive probability that the stack never unwinds back to $\langle i, \ldots, j \rangle$ (this follows from the fact that j is transient and that $\alpha_j < 1$). But if the stack never unwinds to $\langle i, \ldots, j \rangle$, then it never unwinds to $\langle i \rangle$. So there is a positive probability that the stack never unwinds to $\langle i \rangle$, which contradicts the assumption that i is recurrent. Hence, we can assume that for every transient state j, we have $\alpha_j = 1$.

Let j be an arbitrary state. We shall show that j is recurrent, a contradiction. Assume that the backoff process starts in state j; we must show that with probability 1, the stack in the backoff process returns to $\langle j \rangle$. Assume that the next state is ℓ, so that the stack is $\langle j, \ell \rangle$. If ℓ is transient, then with probability 1, on the following step the stack is back to $\langle j \rangle$, since $\alpha_\ell = 1$. Therefore, assume that ℓ is recurrent. So with probability 1, the stack is $\langle j, \ell \rangle$ infinitely often. Since $\alpha_\ell > 0$, it follows that with probability 1, the stack must eventually return to $\langle j \rangle$, which was to be shown.

We have shown that if some state is transient, then they all are. Assume that there is at least one null state and at least one ergodic state; we shall derive a contradiction. This will conclude the proof.

Assume first that there is some null state j with $\alpha_j < 1$. Let i be an ergodic state. There is a positive probability that starting in state i in $(M, \vec{\alpha}, i)$, the backoff process eventually reaches state j and then makes a forward move. Since the expected time in $(M, \vec{\alpha}, j)$ to return to the stack $\langle j \rangle$ is infinite, it follows that the expected time in $(M, \vec{\alpha}, i)$ to return to $\langle i \rangle$ is infinite. This contradicts the assumption that i is ergodic. Hence, for every null state j, we have $\alpha_j = 1$.

Let j be an arbitrary state. We shall show that j is ergodic, a contradiction. For each state i, let h_i be the expected time to return to the stack $\langle i \rangle$ in $(M, \vec{\alpha}, i)$, after starting in state i. Starting in state j in $(M, \vec{\alpha}, j)$, the expected time to return to the stack $\langle j \rangle$ is

$$\sum_{\ell} M_{j\ell}(\alpha_\ell(2) + (1-\alpha_\ell)\alpha_\ell(h_\ell + 2) + (1-\alpha_\ell)^2\alpha_\ell(2h_\ell + 2) + (1-\alpha_\ell)^3\alpha_\ell(3h_\ell + 2) + \cdots) \tag{7}$$

The term $M_{j\ell}\alpha_\ell(2)$ represents the situation where the first step is to some state ℓ followed immediately by a backward step. The term $M_{j\ell}(1-\alpha_\ell)\alpha_\ell(h_\ell + 2)$ represents the situation where the first step is to some state ℓ other than j, followed immediately by a forward step, followed eventually by a return to the stack $\langle j, \ell \rangle$, followed immediately by a backward step. The next term $M_{j\ell}(1-\alpha_\ell)^2\alpha_\ell(2h_\ell + 2)$ represents the situation where the first step is to some state ℓ other than j, followed immediately by a forward step, followed eventually by a return to the stack $\langle j, \ell \rangle$, followed immediately by a forward step, followed eventually by another return to the stack $\langle j, \ell \rangle$, followed immediately by a backward step. The pattern continues in the obvious way.

The contribution to the sum by null states ℓ is finite, since $\alpha_\ell = 1$ for each null state ℓ. Let $z_\ell = h_\ell + 2$. Then

$$(1-\alpha_\ell)\alpha_\ell(h_\ell+2) + (1-\alpha_\ell)^2\alpha_\ell(2h_\ell+2) + (1-\alpha_\ell)^3\alpha_\ell(3h_\ell+2) + \cdots$$

is bounded above by

$$(1-\alpha_\ell)\alpha_\ell(z_\ell) + (1-\alpha_\ell)^2\alpha_\ell(2z_\ell) + (1-\alpha_\ell)^3\alpha_\ell(3z_\ell) + \cdots$$

This is bounded, since

$$(1-\alpha_\ell) + (1-\alpha_\ell)^2(2) + (1-\alpha_\ell)^3(3) + \cdots = (1-\alpha_\ell)/(\alpha_\ell)^2.$$

Therefore, the expression (7), the expected time to return to the stack $\langle j \rangle$, is finite, so j is ergodic, as desired. ■

C.2 Proof of Claim 4.10

Proof [of Claim 4.10]. It is clear by construction that $\gamma > 0$ and thus $\vec{w}$ is admissible. We now show that $\Delta\Phi_{\vec{w}}$ is admissible, by arguing that for every k, the component $\Delta\Phi_{\vec{w},k}$ satisfies the conditions of admissibility.

Case 1: $\alpha_k = 1$. In this case the expected change in potential, $\Delta\Phi_{\vec{w},k}$, is $-1 \cdot w_k = 0$. (Note this is admissible for $\vec{\alpha}$ since $\alpha_k = 1$.)

Case 2: $\alpha_k = 0$. (This includes the case $k = j$.) In this case, we get the following expression for the expected change in potential:

$$\Delta\Phi_{\vec{w},k} = \sum_{k'} M_{kk'} w_{k'}.$$

Since all summands are nonnegative, it suffices to prove one of them is strictly positive. Since $(M, \vec{\alpha})$ is irreducible, we have that there must be some k' such that $M_{kk'} > 0$ and $\alpha_{k'} < 1$. By the latter condition and the admissibility of $\vec{w}$, we get $w_{k'} > 0$ and thus $M_{kk'} w_{k'} > 0$. So $\Delta\Phi_{\vec{w},k} > 0$, as desired.

Case 3: $k \neq j$, $0 < \alpha_k < 1$. Let k' be such that $M_{kk'} > 0$ and $d(k', j) = d(k, j) - 1$. We know such a state k' exists (by definition of shortest paths). We have:

$$\begin{aligned}
\Delta\Phi_{\vec{w},k} &= -\alpha_k \gamma^{d(k,j)} + (1-\alpha_k) \sum_l M_{kl} \gamma^{d(l,j)} \\
&\geq -\alpha_k \gamma^{d(k,j)} + (1-\alpha_k) M_{kk'} \gamma^{d(k',j)} \\
&= \gamma^{d(k',j)} \left(-\alpha_k \gamma + (1-\alpha_k) M_{kk'}\right) \\
&\geq \gamma^{d(k',j)} \left(-\alpha_k \gamma + (1-\alpha_{\max}) M_{\min}\right) \\
&= \gamma^{d(k',j)} \left(-\alpha_k \gamma + 2\gamma\alpha_{\max}\right) \\
&\geq \gamma^{d(k',j)} \left(-\alpha_k \gamma + 2\gamma\alpha_k\right) \\
&= \gamma^{d(k',j)} \left(\alpha_k \gamma\right) \\
&> 0.
\end{aligned}$$

So again, $\Delta\Phi_{\vec{w},k} > 0$, as desired. ■

C.3 Proofs for the classification algorithm

Lemma C.1 *For an $(M, \vec{\alpha})$-backoff process, if there exists an admissible $\vec{w}$ s.t $\Delta\Phi_{\vec{w}}$ is also admissible, then the $(M, \vec{\alpha})$-backoff process is transient.*

Proof. We start by showing that the potential $\Phi_{\vec{w}}(\text{succ}(\text{succ}(\bar{\sigma})))$ has a strictly larger expectation than the potential $\Phi_{\vec{w}}(\bar{\sigma})$. This, coupled with the fact that changes in the potential are always bounded in magnitude, allow us to apply martingale tail inequalities to the sequence $\{\Phi_{\vec{w}}(H_t)\}_t$ and claim that it increases linearly with time, with all but an exponentially vanishing probability. This allows us to prove that with positive probability the walk never returns to the initial history, thus ruling out the possibility that it is recurrent. Details below.

Claim C.2 *There exists an $\epsilon > 0$ s.t. for all sequences $H_0, \ldots, H_t$ of positive probability in the $(M, \vec{\alpha}, i)$-Markov chain,*

$$\mathrm{E}\left[\Phi(H_{t+2}) - \Phi(H_t)\right] > \epsilon.$$

Proof. We start by noticing that the potential must increase (strictly) whenever H_t is the initial history. This is true, since in this case the backoff process is not allowed to backoff. Further, by irreducibility, there exists some state j with $\alpha_j < 1$ and $M_{ij} > 0$. Thus the expected increase in potential from the initial history is at least $\epsilon_1 \stackrel{\text{def}}{=} w_j M_{ij}$. Let ϵ_2 be the smallest non-zero entry of $\Delta\Phi_{\vec{w}}$. We show that the claim holds for $\epsilon = \min\{\epsilon_1, \epsilon_2\}$.

Notice first that both the quantities: $\mathrm{E}\left[\Phi(H_{t+1}) - \Phi(H_t)\right]$ and $\mathrm{E}\left[\Phi(H_{t+2}) - \Phi(H_{t+1})\right]$ are non-negative (since $\Delta\Phi_{\vec{w}}$ is nonnegative). So it suffices to prove that at least one of these quantities increases by at least ϵ. We consider several cases:

Case 1: $\alpha_{\text{top}(H_t)} < 1$: In this case $\mathrm{E}\left[\Phi(H_{t+1}) - \Phi(H_t)\right] = \Delta\Phi_{\vec{w},\text{top}(H_t)} \geq \epsilon_2$, since $\Delta\Phi_{\vec{w}}$ is admissible.

Case 2: $\alpha_{\text{top}(H_t)} = 1$ and $\ell(H_t) > 1$: Let $H_t = \langle \sigma_0, \ldots, \sigma_{l-1}, \sigma_l \rangle$. Note that $H_{t+1} = \langle \sigma_0, \ldots, \sigma_{l-1} \rangle$. Further, note that $\alpha_{\text{top}(H_{t+1})} < 1$ (since only the top or bottom of the history can be states j with $\alpha_j = 1$). Thus, in this case we have, $\mathrm{E}\left[\Phi(H_{t+2}) - \Phi(H_{t+1})\right] \geq \epsilon_2$ (again using the admissibility of $\Delta\Phi_{\vec{w}}$).

Case 3: $\alpha_{\text{top}(H_t)} = 1$ and $\ell(H_t) \leq 1$: In this case, either H_t or H_{t+1} is the initial history, and in such a case, we have that the expected increase in potential is at least ϵ_1. ∎

Next we apply a martingale tail inequality to claim that the probability that the history is the initial history (or equivalently the potential is zero) grows exponentially small with time.

Claim C.3 *There exists $c < \infty$, $\lambda < 1$ such that for every integer $t \geq 0$, the following holds:*

$$\Pr[\ell(H_t) = 0] \leq c \cdot \lambda^t.$$

Proof. Since the potential at the start state is zero, and the potential is expected to go up by ϵ every two time steps, we have that the expected potential at the end of t steps (when t is even) is at least $\epsilon t/2$. Further notice that the sequence $\Phi_{\vec{w}}(H_0), \Phi_{\vec{w}}(H_2), \Phi_{\vec{w}}(H_4), \ldots,$ form a sub-martingale, and that the change in $\Phi_{\vec{w}}(H_t)$ is absolutely bounded: $|\Phi_{\vec{w}}(H_{t+2}) - \Phi_{\vec{w}}(H_t)| \leq 2 \cdot \max_{i \in \{1,\ldots,n\}}\{w_i\}$. Therefore, we can apply a standard tail inequality (Corollary A.5) to show that there exist constants $c < \infty$, $\lambda < 1$ such that

$$\Pr\left[\Phi_{\vec{w}}(H_t) = 0\right] \leq c \cdot \lambda^t.$$

The claim follows by noticing that if the history is the initial history, then the potential is zero. ∎

We use the claim above to notice that for any time T, the probability that the $(M, \vec{\alpha}, i)$-process reaches the initial history after time T is at most $\sum_{t=T}^{\infty} c \cdot \lambda^t \leq c \cdot \lambda^T/(1-\lambda)$. Setting T sufficiently large, we get that this quantity is smaller than 1. Thus the probability that the given $(M, \vec{\alpha}, i)$-backoff process returns to the initial history after time T is bounded away from 1, ruling out the possibility that it is recurrent. ■

Lemma C.4 *For an $(M, \vec{\alpha})$-backoff process, if there exists an admissible $\vec{w}$ s.t the vector $-\Delta\Phi_{\vec{w}}$ is also admissible, then the $(M, \vec{\alpha})$-backoff process is ergodic.*

Proof. First notice that we can modify the vector $\vec{w}$ so that it is positive and $\Delta\Phi_{\vec{w}}$ is negative, as follows: Let ϵ be the smallest non-zero entry of $-\Delta\Phi_{\vec{w}}$. For every j s.t. $\alpha_j = 1$, set $w'_j = w_j + \epsilon/2$. The corresponding difference vector, $\Delta\Phi_{\vec{w}'}$, is at most $\epsilon/2$ larger than $\Delta\Phi_{\vec{w}}$ in any coordinate; and thus entries that were already negative in $\Delta\Phi_{\vec{w}}$ remain negative in $\Delta\Phi_{\vec{w}'}$. On the other hand, for any j such that $\Delta\Phi_{\vec{w},j}$ was 0 (implying $\alpha_j = 1$), the value of $\Delta\Phi_{\vec{w}',j}$ is $-w'_j = -\epsilon/2$. Thus all the zero entries are now negative.

Henceforth we assume, w.l.o.g., that $\vec{w}$ is positive and $\Delta\Phi_{\vec{w}}$ is negative. Let $w_{\min}$ denote the smallest entry of $-\Delta\Phi_{\vec{w}}$ and $w_{\max}$ denote the largest entry of $\vec{w}$. At this stage we have that the expected w-potential always goes down except when the history is an initial history. Notice that when the history is an initial history, the expected increase is potential is at most $w_{\max}$. To deal with initial histories, we define an extended potential.

For a sequence $H_0, \ldots, H_t, \ldots$ of the $(M, \vec{\alpha}, i)$-Markov chain, let $N_0(t)$ denote the number of times the initial history occurs in the sequence $H_0, \ldots, H_{t-1}$. Define the extended potential $\psi(t) = \psi_{\vec{w}}^{H_0,\ldots,H_t,\ldots}(t)$ to be

$$\psi(t) = \Phi_{\vec{w}}(H_t) - (w_{\max} + w_{\min}) \cdot N_0(t).$$

By construction, the extended potential of a sequence is expected to go down by at at least $w_{\min}$ in every step. Thus we have

$$\mathrm{E}[\psi(t)] \leq -w_{\min} \cdot t.$$

Further, the sequence $\psi(0), \ldots, \psi(t), \ldots$ is a super-martingale and the change in one step is absolutely bounded. Thus, by applying a martingale tail inequality (Corollary A.6), we get that for any $\epsilon > 0$, we find that with probability tending to 1, the extended potential after t steps is at most $-(1-\epsilon)w_{\min} \cdot t$. (More formally, $\forall \epsilon, \delta > 0$, there exists a time t_0 such that for every $t \geq t_0$, the probability that the extended potential $\psi(t)$ is greater than $-(1-\epsilon)w_{\min} \cdot t$, is at most δ.) Since the $\Phi_{\vec{w}}$ part of the extended potential is always nonnegative, and each time the sequence reaches the initial history, it is reduced by at most $(w_{\max} + w_{\min})$, this implies that a sequence with potential $-(1-\epsilon)w_{\min} \cdot t$ must include at least $(1-\epsilon)\frac{w_{\min}}{w_{\min}+w_{\max}} \cdot t$ initial histories.

Assume for contradiction that the $(M, \vec{\alpha})$-backoff process is null or transient. Then, the expected time to return to an initial history is infinite. Let Y_i denote the length of the time between the $(i-1)$st and ith visit to the initial history. Using a straightforward application of the law of large numbers (Proposition A.9), we find that for every δ and $c < \infty$ there exists an integer N such that such that with probability at least $1-\delta$, the first N visits to the initial history take more than $c \cdot N$ steps. Setting $\delta = \frac{1}{2}$ and $c = 2 \cdot \frac{w_{\min}+w_{\max}}{(1-\epsilon)w_{\min}}$ and $t = cN$, we find that this contradicts the conclusion of the previous paragraph. We conclude that the $(M, \vec{\alpha})$-backoff process is ergodic. ■

Definition C.5 *For a state j, define the revocation probability as follows: Pick any non-initial history $\bar{\sigma} = \langle \sigma_0, \ldots, \sigma_l \rangle$ with $\mathrm{top}(\bar{\sigma}) = j$. The revocation probability R_j is the probability that the $(M, \vec{\alpha}, i)$-Markov chain starting at state $\bar{\sigma}$ reaches the state $\bar{\sigma}' = \langle \sigma_0, \ldots, \sigma_{l-1} \rangle$. (Notice that this probability is independent of l and $\sigma_0, \ldots, \sigma_{l-1}$; and thus the quantity is well-defined.)*

Before going on to characterize null processes, we prove a simple proposition that we will need in the next lemma.

Proposition C.6 *If an $(M, \vec{\alpha})$-backoff process is transient, then there exists a state j with revocation probability $R_j < 1$.*

Proof. If every state has revocation probability 1, then the first step is revoked with probability 1, indicating that the walk returns to the origin with probability 1, making it recurrent. ■

The converse is also true, but we do not need it, so we do not prove it.

Lemma C.7 *For an $(M, \vec{\alpha})$-backoff process, if there exists an admissible $\vec{w}$ s.t the vector $\Delta\Phi_{\vec{w}} = \vec{0}$ then the $(M, \vec{\alpha})$-backoff process is null.*

Proof. We first define an extended potential as in the proof of Lemma C.4, but we will be a bit more careful. Let $\tau = \mathrm{E}\left[\Phi_{\vec{w}}(H_1) - \Phi_{\vec{w}}(H_0)\right]$ be the expected increase in potential from the initial history. (Note $\tau > 0$.)

For a sequence $H_0, \ldots, H_t, \ldots$ of the $(M, \vec{\alpha}, i)$-Markov chain, let $N_0(t)$ denote the number of occurrences of the initial history in time steps $0, \ldots, t-1$, and let the extended potential $\psi(t)$ be given by

$$\psi(t) = \Phi_{\vec{w}}(H_t) - \tau \cdot N_0(t).$$

Notice that the extended potential is expected to remain unchanged at every step of the backoff process. Applying a martingale tail inequality again (Corollary A.4) we note that for every $\delta > 0$, there exists a constant $c < \infty$ such that the probability that the extended potential $\psi(t)$ is greater than $c\sqrt{t}$ in absolute value is at most δ. We will show that for an ergodic process the extended potential goes down linearly with time, while for a transient process the extended potential goes up linearly with time - thus concluding that the given $(M, \vec{\alpha})$-backoff process fits in neither category.

Claim C.8 *If the $(M, \vec{\alpha})$-backoff process is transient, then there exist constants $\epsilon > 0$ and b s.t. for every time t, it is the case that*

$$\mathrm{E}[\psi(t)] \geq \epsilon t - b.$$

Proof. Let j be a state with $R_j < 1$. Let n be the number of states of the Markov chain M. Notice that for each t and each history H_t, there is a positive probability that there exists a time $t' \in [t+1, t+n]$ such that $\mathrm{top}(H_{t'}) = j$ and the move from $H_{t'-1}$ to $H_{t'}$ is a forward move. Further, conditioned on this event there is a positive probability (of $1 - R_j$) that this move to j is never revoked. Thus in any interval of time of length at least n, there is a positive probability, say γ, that the $(M, \vec{\alpha}, i)$-backoff process makes a move that it never revokes in the future. Thus the expected number of such moves in t steps is $\gamma t/n$. Let $w_{\min}$ be the smallest non-zero entry of $\vec{w}$. Then the expected value of $\Phi_{\vec{w}}(H_t)$ is at least $(\gamma t/n)w_{\min}$.

We now verify that the expected value of $\tau \cdot N_0(t)$ is bounded from above. This is an easy consequence of a well-known property of transient Markov chains, which states that the expected number of returns to the initial state (or any state) is finite. Let this finite bound on $\mathrm{E}[N_0(t)]$ be B. Then for every t, we have $\mathrm{E}[\tau \cdot N_0(t)] \leq \tau B$.

Thus the expected extended potential after t steps is at least $\gamma t/n - \tau B$. ■

Claim C.9 *If the $(M, \vec{\alpha})$-backoff process is ergodic, then there exist constants $\gamma > 0$ and $b < \infty$ such that for all t,*

$$\mathrm{E}\left[\psi(t)\right] \leq -\gamma t + b.$$

Proof. We first argue that the "$-\tau \cdot N_0(t)$" part of the extended potential goes down linearly with time. Let Y_j denote the time between the $(j-1)$st and jth return to the initial history. Then the Y_j's are independently and identically distributed and have a bounded expectation, say T. Then, applying the law of large numbers, we have that there exists a t_0 such that for all $t \geq t_0$ the probability that the number of visits to the origin in the first t time steps is less than $t/2T$ is at most $\frac{1}{2}$. Thus the expected contribution to the extended potential from this part is bounded above by $-\tau \cdot (t-t_0)/(4T)$.

It remains to bound the contribution from $\mathrm{E}[\Phi_{\vec{w}}(H_t)]$. Let $f(t)$ denote the smallest nonnegative index such that the history $H_{t-f(t)}$ is an initial history. Notice then that $\mathrm{E}[\Phi_{\vec{w}}(H_t)]$ is at most $w_{\max} \cdot \mathrm{E}[f(t)]$. We will bound the expected value of $f(t)$. Let $F(t)$ denote this quantity. Let p be the probability distribution on the return time to an initial history, starting from H_0. Recall that $\sum_i ip(i) = T$. Then $F(t)$ satisfies the relation:

$$F(t) = \sum_{i=1}^{t} p(i)F(t-i) + \sum_{i=t+1}^{\infty} tp(i).$$

(If the first return to the initial history happens at time i and $i > t$, then $f(t) = t$, and if $i \leq t$ then $f(t) = f(t-i)$.) We use this relation to prove, by induction on t, that: For every $\epsilon > 0$, there exists a constant a such that $F(t) \leq \epsilon t + a$. Set a such that $\sum_{i>a} ip(i) \leq \frac{\epsilon}{2}T$. The base cases of the induction are with $t \leq a$ and these easily satisfy the hypothesis, since $F(t) \leq t \leq a \leq \epsilon t + a$. For $t > a$, we get:

$$\begin{aligned}
F(t) &= \sum_{i=1}^{t} p(i)F(t-i) + \sum_{i=t+1}^{\infty} tp(i) \\
&\leq \sum_{i=1}^{t} p(i)(\epsilon(t-i)+a) + \sum_{i=t+1}^{\infty} tp(i) \\
&\leq \sum_{i=1}^{\infty} p(i)\epsilon t - \sum_{i=1}^{t} p(i)\epsilon i + \sum_{i=1}^{\infty} p(i)a + \sum_{i=t+1}^{\infty} ip(i) \\
&= \epsilon t + a - \sum_{i=1}^{\infty} p(i)\epsilon i + \sum_{i=t+1}^{\infty} (1+\epsilon)ip(i) \\
&\leq \epsilon t + a - \epsilon T + (1+\epsilon)(\epsilon/2)T \\
&\leq \epsilon t + a \text{ (Using } \epsilon \leq 1).
\end{aligned}$$

The claim now follows by setting $\epsilon = \frac{\tau}{8T}$ and $b = \frac{\tau \cdot t_0}{8T} + l$. ■

■

C.4 Existence of Cesaro Limits

In this section we prove that the $(M, \vec{\alpha}, i)$-backoff process always converges to a Cesaro limit. The proofs are different for each case (ergodic, null and transient), and so we divide the discussion based on the case.

C.4.1 Ergodic case

The simplest argument is for the ergodic case.

Theorem C.10 *If the backoff process $(M, \vec{\alpha})$ is ergodic, then the $(M, \vec{\alpha}, i)$-backoff process converges to a Cesaro limit.*

Proof. We gave this proof in Section 4.2. ■

C.4.2 Transient case

In this section, we consider the transient case (where the Polish matrix is transient).

Theorem C.11 *If the $(M, \vec{\alpha})$-backoff process is transient, then the $(M, \vec{\alpha}, i)$-backoff process converges to a stationary distribution.*

Proof. Since the Polish matrix is transient, we know that for each state $\bar{\sigma}$ of the Polish matrix (which is a stack of states of the backoff process) where the top state $\mathrm{top}(\bar{\sigma})$ has $\alpha_{\mathrm{top}(\bar{\sigma})} \neq 1$, there is a positive probability, starting in $\bar{\sigma}$, that the top state $\mathrm{top}(\bar{\sigma})$ is never popped off of the stack. It is clear that this probability depends only on the top state $\mathrm{top}(\bar{\sigma})$ of the stack $\bar{\sigma}$.

When the backoff process is in a state (with a given stack), and that state is never popped off of the stack (by taking a backedge), then we refer to this (occurrence of the) state as *irrevocable*. Technically, an irrevocable state should really be thought of as a pair consisting of the state (of the backoff process) and the time, but for convenience we shall simply refer to the state itself as being irrevocable.

We now define a new matrix, which we call the *Turkish matrix*, which defines a Markov chain. Just as with the Polish matrix, the states are again stacks of states of the backoff process, but the interpretation of the stack is different from that of the Polish matrix. In the Turkish matrix, the stack $\langle \sigma_0, \ldots, \sigma_\ell \rangle$ represents a situation where σ_0 is irrevocable, and where $\sigma_1, \ldots, \sigma_\ell$ are *not* irrevocable. The intuition behind the state $\langle \sigma_0, \ldots, \sigma_\ell \rangle$ is that the top states of the stack of the Turkish matrix (from σ_0 on up) are $\sigma_0, \ldots, \sigma_\ell$. As with the Polish matrix, the states $\langle \sigma_0, \ldots, \sigma_\ell \rangle$ of the Turkish matrix are restricted to being the attainable ones: in this case this means (a) $\alpha_{\sigma_j} \neq 1$ for $0 \leq j < \ell$; (b) $\alpha_{\sigma_j} \neq 0$ for $1 \leq j \leq \ell$; and (c) $M_{\sigma_i \sigma_{i+1}} > 0$ for $0 \leq i < \ell$. There is a subtlety if the start state i has $\alpha_i = 1$, since then the state $\langle i \rangle$ is not reachable from any other state, and so we do not consider it to be a state of the Turkish matrix. One way around this issue is simply to assume that the start state i has $\alpha_i \neq 1$. This is an acceptable assumption, since with probability 1, the backoff process will reach a state j with $\alpha_j \neq 1$ in a finite number of steps, and ignoring a finite number of steps has no effect on asymptotic probabilities.

We now define the entries of the Turkish matrix T. If $\bar{\sigma}$ and $\bar{\sigma}'$ are states of the Turkish matrix, then the entry $T_{\bar{\sigma}\bar{\sigma}'}$ is 0 unless either (a) $\bar{\sigma}'$ is the result of popping the top element off of the stack $\bar{\sigma}$, (b) $\bar{\sigma}'$ is the result of pushing one new element onto the stack $\bar{\sigma}$, or (c) both $\bar{\sigma}$ and $\bar{\sigma}'$ each contain exactly one element. The probabilities are those induced by the backoff process. Thus, in case (a), if $\ell \geq 1$, then $T_{\langle \sigma_0, \ldots, \sigma_\ell \rangle \langle \sigma_0, \ldots, \sigma_{\ell-1} \rangle}$ equals the probability that the backoff process takes a backedge from σ_ℓ, given that the last irrevocable state was σ_0, that the stack from σ_0 on up is $\langle \sigma_0, \ldots, \sigma_\ell \rangle$, and that the remaining states $\sigma_1, \ldots, \sigma_{\ell-1}$ on the stack are not irrevocable. That this conditional probability is well-defined (and is independent of the time) can be seen by writing $\Pr[A \mid B]$ as $\Pr[A \wedge B] / \Pr[B]$. Note that even though this conditional probability represents the probability of taking a backedge from state σ_ℓ, it is not necessarily equal to α_{σ_ℓ}, since the event of taking the backedge is conditioned on other events, such as that σ_0 is irrevocable. Similarly, in case (b), we have that $T_{\langle \sigma_0, \ldots, \sigma_\ell \rangle \langle \sigma_0, \ldots, \sigma_{\ell+1} \rangle}$ equals the probability that the backoff process takes a forward edge from σ_ℓ to $\sigma_{\ell+1}$ and that $\sigma_{\ell+1}$ is not irrevocable, given that the last irrevocable state was σ_0, that the stack from σ_0 on up is $\langle \sigma_0, \ldots, \sigma_\ell \rangle$, and that the remaining states $\sigma_1, \ldots, \sigma_\ell$ on the stack are not irrevocable. Finally, in case (c) we have that $T_{\langle\sigma_0\rangle\langle\sigma_0'\rangle}$ equals the probability that the backoff process takes a forward edge from to σ_0 to σ_0' and that σ_0' is irrevocable, given that σ_0 is irrevocable.

We now show that the Turkish matrix is irreducible, aperiodic, and (most importantly) ergodic.

We first show that it is irreducible. We begin by showing that from every state of the Turkish matrix, it is possible to eventually reach each (legal) state $\langle\sigma_0\rangle$ with only one element in the stack (by "legal", we mean that $\alpha_{\sigma_0} \neq 1$). This is because in the backoff process, it is possible to eventually reach the state σ_0, because the backoff process is irreducible; further, it is possible that once this state σ_0 is reached, it is then irrevocable. Next, from the state $\langle\sigma_0\rangle$, it is possible to eventually reach each state $\langle\sigma_0,\ldots,\sigma_\ell\rangle$ with bottom element σ_0. This is because it is possible to take forward steps from σ_0 to σ_1, then to σ_2, ..., and then to σ_ℓ, with each of the states $\sigma_1,\sigma_2,\ldots,\sigma_\ell$ being non-irrevocable (they can be non-irrevocable, since it is possible to backup from σ_ℓ to $\sigma_{\ell-1}$ ... to σ_0). Combining what we have shown, it follows that the Turkish matrix is irreducible.

We now show that the Turkish matrix is aperiodic. Let i be a state with $\alpha_i \neq 1$. Since the backoff process is aperiodic, the gcd of the lengths of all paths from i to itself is 1. But every path from i to itself of length k in the backoff process gives a path from $\langle i\rangle$ to itself of length k in the Turkish matrix (where we take the arrival in state i at the end of the path to be an irrevocable state). So the Turkish matrix is aperiodic.

We now show that the Turkish matrix is ergodic. It is sufficient to show that for some state of the Turkish matrix, the expected time to return to this state from itself is finite. We first show that the expected time between irrevocable states is finite. Thus, we shall show that the expected time, starting in an irrevocable state σ_0 in the backoff process at time t_0, to reach another irrevocable state is finite. Let E_k be the event that the time to reach the next irrevocable state is at least k steps (that is, takes place at time t_0+k or later). It is sufficient to show that the probability of E_k is $O(\lambda^k)$ for some constant $\lambda<1$. Assume that the event E_k holds. There are now two possible cases. *Case 1:* There are no further irrevocable states. In this case the state of the Turkish matrix is of the form $\langle\sigma_0,\sigma_1\rangle$ infinitely often with probability 1. *Case 2:* There is another irrevocable state, that occurs at time t_0+k or later. Assume that it occurs for the first time at time t_0+k', where $k' \geq k$. It is easy to see that the state of the Turkish matrix at time t_0+k' is of the form $\langle\sigma_0,\sigma_1\rangle$. So in both cases, there is $k' \geq k$ such that after k' steps, the size of the stack in the Polish matrix has grown by only one.

Now since the Polish matrix is transient, we see from Section 4.1 that we can define a potential such that there is an expected positive increase in the potential at each step. So by a submartingale argument (Corollary A.5), the probability that the size of the stack in the Polish matrix has grown by only one after $k'>k$ steps is $O(\lambda^{k'})$ for some constant $\lambda<1$. So the probability of E_k is $O(\lambda^k)$, as desired.

We have shown that the expected time between irrevocable states is finite. So starting in $\langle\sigma_0\rangle$, there is some state σ_1 such that the expected time to reach $\langle\sigma_1\rangle$ from $\langle\sigma_0\rangle$ is finite. Continuing, we see that there is some state σ_2 such that the expected time to reach $\langle\sigma_2\rangle$ from $\langle\sigma_1\rangle$ is finite. Similarly, there is some state σ_3 such that the expected time to reach $\langle\sigma_3\rangle$ from $\langle\sigma_2\rangle$ is finite, and so on. Let n be the number of states in the backoff process. Then some state σ appears at least twice among $\sigma_0,\sigma_1,\ldots,\sigma_n$. Hence, the expected time from $\langle\sigma\rangle$ to itself in the Turkish matrix is finite. This was to be shown.

We have shown that the Turkish matrix is irreducible, aperiodic, and ergodic. So it has a steady-state distribution. This gives us a stedy-state distribution in the backoff process, where the probability of state i is the sum of the probabilities of the stacks in the Turkish matrix with top state i. ■

C.4.3 Null case

I like the use of the work "walk" in this section. Can we make this legal?

In this section we prove the following theorem.

Theorem C.12 *The $(M, \vec{\alpha}, i)$ process has a Cesaro limit if the $(M, \vec{\alpha})$ process is null.*

The theorem is implied by the lemma below.

Lemma C.13 *Let $(M, \vec{\alpha})$ be null. Let j be any state of M such that $\alpha_j < 1$. Let $\vec{\alpha}'$ be the vector given by $\alpha'_j = 1$ and $\alpha'_{j'} = \alpha_{j'}$ otherwise. Then $(M, \vec{\alpha}')$ is ergodic and hence has a Cesaro limit distribution. Let i be any state of M. Then the $(M, \vec{\alpha}, i)$ process has a Cesaro limiting distribution which is the same as the Cesaro limit distribution of the $(M, \vec{\alpha}', j)$ process.*

Remark: There are some magical consequences to the above lemma.

Proof. The first part of Lemma C.13 claiming that $(M, \vec{\alpha}')$ is ergodic, follows from a simple monotonicity argument, proven in Claim C.16. We now move to the more hairy part. For this part, we consider a walk W of length t of the $(M, \vec{\alpha}, i)$ process and break it down into a number of smaller pieces. This breakdown is achieved by a "skeletal decomposition" as defined below.

Fix an $(M, \vec{\alpha}, i)$ walk W with $\langle X_0, \ldots, X_t \rangle$ being the sequence of states visited, with auxiliary sequence $\langle S_0, \ldots, S_t \rangle$ and associated history sequence $\langle H_0, \ldots, H_t \rangle$.

For every $t_1 \leq t$ such that $S_{t_1} =$"F" (i.e., W makes a forward step at time t_1), we define a partition of W into two walks W' and W'' as follows: Let j be the state pushed onto the history stack at time t_1 and let $H_{t_1} = \bar{\sigma}$ be the history stack at time t_1. Let $t_2 > t_1$ be the first time at which this history repeats itself ($t_2 = t$ if this event never happens). Consider the sequence of time steps $\langle 0, \ldots, t_1, t_2 + 1, \ldots, t \rangle$ (and the associated sequence of states visited and auxiliary sequences). They give a new $(M, \vec{\alpha}, i)$ walk W' that has positive probability. On the other hand the sequence of time steps $t_1, t_1 + 1, \ldots, t_2$ define a walk W'' of an $(M, \vec{\alpha}, j)$ process, of length $t_2 - t_1$, with initial history being $\langle j \rangle$. We call this partition (W', W'') a *j-division* of the walk W. (Notice that W', W'' do not suffice to recover W, and this is fine by us.) A *j-decomposition* of a walk W is an (unordered) collection of walks $W_0, \ldots, W_k$ that are obtained by a sequence of j-decompositions of W. Specifically, W is a j-decomposition of itself; and if $W_0, \ldots, W_l$ is a j-decomposition of W'; $W_{l+1}, \ldots, W_k$ is a j-decomposition of W''; and W', W'' is a j-division of W, then $W_0, \ldots, W_k$ is a j-decomposition of W. If a walk has no non-trivial j-divisions, then it is said to be j-indivisible. A j-skeletal decomposition of a walk W is a j-decomposition $W_0, \ldots, W_k$ of W, where each W_l is j-indivisible. Note that the skeletal decomposition is unique and independent of the choice of j-divisions. We refer to $W_0, \ldots, W_k$ as the skeletons of W. Note that the skeletons come in one of three categories (assuming $j \neq i$).

- Initial skeleton: This is a skeleton that has $\langle i \rangle$ as its initial history. Note that there is exactly one such skeleton. (If $i = j$, we say there are no initial skeletons.)
- Closed skeletons: These are the skeletons with $\langle j \rangle$ as their initial and final history.
- Open skeletons: These are the skeletons with $\langle j \rangle$ as their initial, but not their final history.

Our strategy for analyzing the frequency of the occurence of a state j' in the walk W is to decompose W into its skeletons and then to examine the relative frequency of j' in these skeletons. Roughly we will show that not too much time is spent in the initial and open skeletons; and that the distribution of closed skeletons of W is approximated by the distribution of random walks returning to the initial history in an $(M, \vec{\alpha}', j)$-backoff process. But the $(M, \vec{\alpha}', j)$ process is ergodic and thus the expected time to return to the origin in such walks is finite and thus in a large number of closed j-skeletons, the frequency of occurence of j' converges (to its frequency in $(M, \vec{\alpha}', j)$-processes).

Simulation of W.

1. Pick an (infinite) walk W_0' from the $(M, \vec{\alpha}', i)$ process.
2. Pick a sequence of walks $W_1', W_2', \ldots,$ as follows: For each k, W_k' starts at $\langle j \rangle$ and walks according to $(M, \vec{\alpha}', j)$ and terminates the first time it returns to the initial history.
3. We now cut and paste from the W_i''s to get W as follows:
   (a) We initialize $W = W_0'$ and $t' = 0$, $N = 0$.
   (b) We iterate the following steps till $t' \geq t$:
      i. Let t'' be the first visit to j occurring at some time after t' in W. Set $t' = t''$.
      ii. With probability α_j do nothing, else (with probability $1 - \alpha_j$, set $N = N + 1$ and splice the walk W at time t' and insert the walk W_N' into W at this time.
   (c) Truncate W to its first t steps and output it. Further, let W_i denote the truncation of W_i' up to the point to which it is used in W.

The following proposition is easy to verify.

Proposition C.14 *W generated as above has exactly the same distribution as that of the random $(M, \vec{\alpha}, i)$-process. Further $W_0, \ldots, W_N$ give the j-skeletal decomposition of W.*

Let W' denote a random walk obtained by starting at $\langle j \rangle$, walking according to $(M, \vec{\alpha}', j)$ and stopping the first time we reach the initial history. Since the $(M, \vec{\alpha}', j)$ process is ergodic, the expected length of W' is finite. Let μ denote the expectation of the length of the walk W' and let $\mu_{j'}$ denote the expected number of occurrences of the state j' in W'. Note that $\mu_{j'}/\mu = \pi'_{j'}$, where π' denotes the stationary distribution of the $(M, \vec{\alpha}', j)$ process.

Let a_k' denote the number of visits to j' in W_k' and let b_k' denote the length of W_k'. Since the walks W_k' $(k \in \{1, \ldots, N\})$ are chosen independently from the same distribution as W', we have that the expectation of a_k' is $\mu_{j'}$ and the expectation of b_k' is μ. Let a_k denote the number of visits to j' in W_k and let b_k denote the length of W_k. Notice our goal is to show that $\sum_{k=0}^{N} a_k / \sum_{k=0}^{N} b_k$ approaches $\pi_{j'}$ with probability tending to 1 as t tends to infinity. Fix any $\beta > 0$. We will enumerate a number of bad events, argue that each one of them has low probability of occurence and then argue that if none of them happen, then

$$(1 - \beta)\pi_{j'} \leq \sum_{k=0}^{N} a_k / \sum_{k=0}^{N} b_k \leq (1 + \beta)\pi_{j'},$$

1. N is too small: In Claim C.17 we show that this event has low probability. Specifically, there exists a $\delta > 0$ such that for every $\epsilon > 0$ there exists a t_0 such that for all $t \geq t_0$, the probability that N is less that δt is at most ϵ.
2. W_0 is too long: Claim C.18 shows that for every $\epsilon > 0$, there exists t_1 such that for all $t \geq t_1$, the probability that W_0 is longer than ϵt is at most ϵ.

3. There are too many open skeletons: In Claim C.20, we prove that for every $\epsilon_0 > 0$, there exists an t_2 such that if $t \geq t_2$, then the probability that the number of open skeletons is more than $\epsilon_0 t$ is at most ϵ_0.

4. $\sum_{k=1}^{N} b_k$ is too large: By the law of large numbers, we have that for every $\epsilon > 0$, there exists N_1 such that for all $N \geq N_1$, the probability that $\sum_{k=1}^{N} b'_k \geq (1+\epsilon)\mu N$ is at most ϵ. Using the fact that $b_k \leq b'_k$, we obtain the same upper bound on $\sum_k b_k$ as well.

5. $\sum_{k=1}^{N} a_k$ is too large: As above, we have that we have that for every $\epsilon > 0$, there exists N_2 such that for all $N \geq N_2$, the probability that $\sum_{k=1}^{N} a_k \geq (1+\epsilon)\mu_{j'} N$ is at most ϵ.

6. (Informally) $\sum_{k=1}^{N} b_k$ is too small: The formal event considered here is that for some large subset $S \subseteq \{1, \ldots, N\}$, the quantity $\sum_{k \in S} b'_k$ turns out to be too small. Using the fact that the b'_k's are independently and identically distributed and have finite mean μ, Claim C.21 can be used to show that for every $\epsilon > 0$, there exists an $\epsilon_1 > 0$ and $N_3 > 0$, such that for all $N \geq N_3$ the probability that there exists a subset $S \subseteq \{1, \ldots, N\}$ of cardinality at least $(1-\epsilon_1)N$ such that $\sum_{k \in S} b'_k \leq (1-\epsilon)\mu N$ is at most ϵ. Taking S to be the subset of closed skeletons and using the fact that for a closed skeleton $b_k = b'_k$, and relying on Item (3), we get to the informal claim here.

7. $\sum_{k=1}^{N} a_k$ is too large: Obtained as above. Specifically, for every $\epsilon > 0$, there exists an $\epsilon_2 > 0$ and $N_4 > 0$, such that for all $N \geq N_4$ the probability that there exists a subset $S \subseteq \{1, \ldots, N\}$ of cardinality at least $(1-\epsilon_2)N$ such that $\sum_{k \in S} b'_k \leq (1-\epsilon)\mu N$ is at most ϵ.

Given the above claims, the lemma may be proved as follows: Let δ be as in Item (1) above. Given any β, let $\epsilon = \min\{\beta/7, \beta/(2+1/(\mu\delta)), \beta/(2+1/(\mu_{j'}\delta)+\beta)\}$. Let ϵ_1 and ϵ_2 be as given in Items (6) and (7) above and let $\epsilon_0 = \min\{\epsilon, \epsilon_1\delta, \epsilon_2\delta\}$. For these choices of ϵ and ϵ_0, let $t_0, t_1, t_2, N_1, N_2, N_3, N_4$ be as given in Items (1)-(7) and let $t \geq \max\{t_0, t_1, t_2, \frac{1}{\delta}N_1, \frac{1}{\delta}N_2, \frac{1}{\delta}N_3, \frac{1}{\delta}N_4\}$. Then since t is large enough, we have that for any of Items (1), (2), or (3) the probability of the bad event listed there happens is at most ϵ. If the bad event of Item (1) does not occur, then $N \geq \{N_1, N_2, N_3, N_4\}$ and thus the probability of any of the bad events list in Items (3)-(7) is at most ϵ. Summing over all bad events, we have the probability that no bad events happens is at least $1 - 7\epsilon \geq 1 - \beta$. We now reason that if none of these events happen then $\frac{\sum_{k=0}^{N} a_k}{\sum_{k=0}^{N} b_k}$ is between $(1-\beta)\pi'_{j'}$ and $(1+\beta)\pi'_{j'}$. We show the lower upper bound. The upper bound is similar. We first upper bound $\sum_{k=0}^{N} b_k$ by Items (2) and (4). By Item (2) $b_0 \leq \epsilon t \leq \frac{\epsilon}{\delta} N$ (where the second inequality uses Item (1).) By Item (4) $\sum_{k=1}^{N} b_k \leq (1+\epsilon)\mu N$ and thus we have

$$\sum_{k=0}^{N} b_k \leq (1 + \epsilon + \epsilon/(\mu\delta))\mu N.$$

Next to lower bound $\sum_{k=0}^{N} a_k$, we use Item (3) to conclude that the number of closed skeletons is at least $N - \epsilon_0 t \geq N - (\epsilon_0/\delta)N \geq (1-\epsilon_2)N$. Let S denote the collection of closed skeletons. Thus, we have

$$\sum_{k=0}^{N} a_k \geq \sum_{k \in S} a_k = \sum_{k \in S} a'_k \geq (1-\epsilon)\mu_{j'} N.$$

Putting the above together, we get

$$\frac{\sum_{k=0}^{N} a_k}{\sum_{k=0}^{N} b_k} \geq \frac{1-\epsilon}{1+\epsilon+\epsilon/(\mu\delta)} \frac{\mu_{j'}}{\mu} \geq (1-\beta)\pi'_{j'},$$

as desired. (The final inequality above uses $\pi'_{j'} = \mu_{j'}/\mu$ and $\epsilon \le \beta/(2+1/(\mu\delta))$.) The upper bound follows similarly, using the inequality $\epsilon \le \beta/(2+1/(\mu_{j'}\delta)+\beta)$. This concludes the proof of the lemma, modulo Claims C.16-C.21. ■

For the following claims, let H denote the Hungarian matrix corresponding to the $(M, \vec{\alpha})$ process and let H' denote the Hungarian matrix corresponding to the $(M, \vec{\alpha}')$ process. For a non-negative matrix A, let $\rho(A)$ denote its maximal eigenvalue. For $n \times n$ matrices A and B, say $A < B$ if $A_{ik} \le B_{ik}$ for every i, k and there exists i, k such that $A_{ik} < B_{ik}$. Claim C.16 will use the following simple claim.

Claim C.15 *If A and B are $n \times n$ irreducible non-negative matrices such that $A < B$, then $\rho(A) < \rho(B)$.*

Proof. Notice first that it suffices to prove that $\rho(I+A) < \rho(I+B)$, since $\rho(I+M) = 1 + \rho(M)$. Similarly it suffices to prove that for some postive integer k. $\rho((I+A)^k) = \rho((I+B)^k)$, since $\rho(M^k) = \rho(M)^k$. We will do so for $k = 2n-1$. Let $C = (I+A)^{2n-1}$ and $D = (I+B)^{2n-1}$.

We first show that for every pair i, j, $C_{ij} < D_{ij}$. Notice that the i,jth entry of a matrix M^k has the following combinatorial interpretation: It counts the sum of the weights of all walks of length k between i and j where the weights of a walk is the product of the weight of the edges it takes, and the weight of an edge (u, v) is M_{uv}. Thus we wish to show that for every i, j, there exists a walk P from i to j of length $2n-1$ such that its weight under $I + A$ is less than its weight under j. Let $A_{lm} < B_{lm}$. By irreducibility of A we know there exists a path from i to l of positive weight and by taking enough self-loops this can be converted into a path P_1 of length exactly $n-1$ with positive weight in $(I+A)$. The path has at least the same weight in $I+B$. Similarly we can find a path P_2 of positive weight in $I+A$ from m to j of length exactly $n-1$. Now the path $P_1 \circ (l, m) \circ P_2$ has positive weight in both $I+A$ and $I+B$ and has strictly larger weight in $I+B$ since $B_{lm} > A_{lm}$. Thus we find that $C_{ij} < D_{ij}$, for every pair i, j.

Now we use the properties of the maximal eigenvalue to show that $\rho(C) < \rho(D)$. Notice that

$$\rho(C) = \max_{\vec{x}} \min_{i \in \{1,\ldots,n\}} \left\{ \frac{(C\vec{x})_i}{(\vec{x})_i} \right\}.$$

Pick $\vec{x}$ that maximizes the right hand side above and now consider

$$\begin{aligned} \rho(D) &= \max_{\vec{y}} \min_{i \in \{1,\ldots,n\}} \left\{ \frac{(D\vec{y})_i}{(\vec{y})_i} \right\} \\ &\ge \min_{i \in \{1,\ldots,n\}} \left\{ \frac{(D\vec{x})_i}{(\vec{x})_i} \right\} \\ &> \min_{i \in \{1,\ldots,n\}} \left\{ \frac{(C\vec{x})_i}{(\vec{x})_i} \right\} \text{(Since } D_{ij} > C_{ij} \text{ and } \vec{x} \neq 0\text{)}. \\ &= \rho(C) \text{(By our choice of } \vec{x}.\text{)} \end{aligned}$$

■

We are now ready to prove that the $(M, \vec{\alpha}')$ process is ergodic.

Claim C.16 *Let $(M, \vec{\alpha})$ be irreducible and null. Let j be a state such that $\alpha_j < 1$. Let $\vec{\alpha}'$ be the vector obtained by setting $\alpha'_{j'} = \alpha_{j'}$ if $j' \neq j$ and $\alpha'_j > \alpha_j$. Then $(M, \vec{\alpha}')$ is ergodic (though it may not be irreducible).*

Proof. We first focus on the case $\alpha'_j < 1$. In this case, we observe that $(M, \vec{\alpha}')$ is also irreducible. For this part, we use the proof of Theorem 4.2 to rephrase this question in terms of the maximal eigenvalues of the corresponding Hungarian matrices. In particular, we have $\rho(H) = 1$ and we need $\rho(H') < \rho(H) = 1$.

Note that for every k, l, we have

$$\begin{array}{rcl} H'_{kl} & = & (1-\alpha'_k) M_{kl} \alpha'^{-1}_l \\ & \leq & (1-\alpha'_k) M_{kl} \alpha_l^{-1} \\ & \leq & (1-\alpha_k) M_{kl} \alpha_l^{-1} \\ & = & H_{kl} \end{array}$$

Further, the first inequality is strict if $l = j$ and $M_{kj} \neq 0$ (and such a k does exist, by the irreducibility of M). Using Claim C.15 we now have $\rho(H') < \rho(H) = 1$. and thus we have shown the desired result for the case $\alpha'_j < 1$.

For the case $\alpha'_j = 1$, we first use the first part shown above to show that the $(M, \vec{\alpha}'')$ process, where $\alpha_j < \alpha''_j < 1$ (and $\alpha''_{j'} = \alpha_{j'}$ for other j'), is ergodic. Thus it suffices to prove that $(M, \vec{\alpha}')$ is ergodic, given that $(M, \vec{\alpha}'')$ is ergodic. However, since we may not have irreducibility, we need to argue this individually for every $(M, \vec{\alpha}', i)$ process. We will do so by arguing that the expected return time of an $(M, \vec{\alpha}', i)$ process (to its initial history) is finite. We use the fact that the expected return time of the $(M, \vec{\alpha}'', i)$ process is finite.

Given a walk W of the $(M, \vec{\alpha}'', i)$ process, let $I(W)$ denote the initial skeleton W_0 in the j-skeletal decomposition of W. Let $S(W_0)$ denote the set of walks W such that $I(W) = W_0$. Let $p(W)$ denote the probability of the walk W in the $(M, \vec{\alpha}'', i)$ process; and let $p'(W_0)$ denote the probability of the walk W_0 in the $(M, \vec{\alpha}', i)$ process. Notice that $\sum_{W \in S(W_0)} p(W) = p'(W_0)$ and the length of W_0 is at most the length of W for every $W \in S(W_0)$. Putting these together, we find the expected length of W_0 in the $(M, \vec{\alpha}', i)$ process is at most the expected length of W in the $(M, \vec{\alpha}'', i)$ process. ∎

The next claim shows that N, the number of skeletons in a walk of length t, grows linearly in t.

Claim C.17 *There exists a $\delta > 0$, such that for every $\epsilon > 0$ there exists a t_0 such that for all $t \geq t_0$, the probability that N is less than δt is at most ϵ.*

Proof. Notice that the number of skeletons is lower bounded by the number of times j is pushed on to the history stack in the walk W. We lower bound this quantity by using the fact that in any sequence of n steps (where n is the size of the Markov chain M), there is a positive probability ρ of pushing j onto the history stack in the next n steps. Thus the expected number of times j is pushed onto the history in t steps is at least $\rho(t/n)$. Applying the law of large numbers, we get that the there exists t_0 s.t. if $t \geq t_0$, then the probability that j is pushed on the stack fewer than $\frac{1}{2}\rho(t/n)$ times is at most ϵ. The claim follows with $\delta = \frac{\rho}{2n}$. ∎

Next we argue that the initial skeleton is not too long.

Claim C.18 *For every $\epsilon > 0$, there exists a time t_1 such that for all time $t > t_1$,*

$$\Pr[\textit{ Length of } W_0 > \epsilon t] < \epsilon.$$

Proof. We prove the claim in two steps. First we note that in a walk of length t, with high probability, the (null) $(M, \vec{\alpha}, i)$ process returns to the initial history $o(t)$ times. Note that the expected time to return to the initial history is infinite. Thus we get:
Sub claim 1: For every $\epsilon' > 0$, there exists a time t'_1 such that for all $t > t'_1$, the probability that an $(M, \vec{\alpha}, i)$ walk of length t returns to the initial history more than $\epsilon' t$ times is at most ϵ'.

We then note that $o(t)$ returns to the initial history of the (ergodic) $(M, \vec{\alpha}', i)$ process are unlikely to take ϵt time steps.
Sub claim 2: Let T be the expected return time to the initial history in the $(M, \vec{\alpha}', i)$ process. (Note $T < \infty$.) Then, for every ϵ'', there exists an N_0 such that if $N \geq N_0$ and $N' \leq N$, then the probability that N' returns to the origin take more than $2NT$ steps is at most ϵ''.

¿From the two sub-claims, we get the claim as follows: Set $\epsilon'' = \epsilon/2$ and $\epsilon' = \min\{\epsilon/2, \epsilon/(2T)\}$. Now let N_0 and T be as in Sub claim 2 and let $t_0 = \max\{t'_0, \frac{2N_0T}{\epsilon}\}$. Given $t \geq t_0$, let $N = (\epsilon t)/(2T)$. Notice $N \geq N_0$. Let N' denote the number of returns to the initial history in W. Applying Sub claim 1 with ϵ' we get that probability that the number of returns to the initial history is more than N is at most $\epsilon' \leq \epsilon/2$. Now applying Sub claim 2 with parameter ϵ'', we get that the probability of N returns to the origin taking more that $2NT = \epsilon t$ steps is at most $\epsilon'' = \epsilon/2$. Thus the probability that any of bad events listed in the Sub claims above occur is at most ϵ, and if neither occurs, then the length of the initial skeleton is at most ϵt. ■

Next we show that not too many skeletons are open. We do it in two claims.

Claim C.19 *If $(M, \vec{\alpha}, i)$ is null, and $\vec{w}$ is a weight vector as guaranteed to exist by Lemma 4.9, then the $\vec{w}$-potential $\Phi_{\vec{w}}(H_t)$ is expected to grow as $o(t)$.*

Proof. Recall the extended potential used in Lemma C.4 was expected to be 0 after t steps. Further, by Sub claim 1 of Claim C.18, the number of returns to the initial history is at most $\epsilon' t$, with probability all but ϵ'. Thus the expected number of returns to the origin is at most $2\epsilon' t$. Thus the expected value of $\phi_{\vec{w}}(H_t)$ is also at most $2\epsilon' t$. ■

Claim C.20 *For every $\epsilon > 0$, there exists a t_2 such that for all $t \geq t_2$. the probability that more than ϵt of the skeletons $W_1, \ldots, W_N$ are open is at most ϵ.*

Proof. Consider the event E that the history H_t contains more than ϵt occurences of the state j. We wish to show that the probability that E occurs is at most ϵ. Assume E occurs with probability at least ϵ. Let $\vec{w}$ be the weight vector as shown to exist in Lemma 4.9, and let $\phi_{\vec{w}}(H_t)$ be the potential of the history H_t. Notice that if E occurs, then the potential $\phi_{\vec{w}}(H_t)$ is at least $w_j \epsilon t$. Since E happens with probability at least ϵ, we have that the expected potential, $\mathrm{E}[\phi(H_t)] \geq \epsilon^2 w_j t$, i.e., it is growing linearly in t. But this contradicts the previous claim. ■

Finally we conclude with a technical claim showing that large subsets of $\{1, \ldots, N\}$ can not have a small sum.

Lemma C.21 *For every distribution $\mathcal{D}$ on non-negative integers with finite expectation μ, and every $\epsilon > 0$, there exists an $\epsilon_1 > 0$ and $N_3 > 0$, such that such that for all $N \geq N_3$, if $X_1, \ldots, X_N$ are N samples drawn i.i.d. from $\mathcal{D}$, then*

$$\Pr\left[\forall S \subseteq [N], |S| \geq (1-\epsilon_1)N, \quad \sum_{i \in S} X_i \geq (1-\epsilon)\mu N\right] \geq 1 - \epsilon.$$

Proof. We will pick τ such that with high probability the ϵN-th largest element of $X_1, \ldots, X_N$ is greater than or equal to τ. We will then sum only those elements in the X_i's whose value is at most τ and this will give a lower bound on $\sum_{i \in S} X_i$.

Let $p(j)$ be the probability given to j by $\mathcal{D}$. Let $\mu_k = \sum_{j \le k} jp(j)$. Notice μ_k's converge to μ. Let τ be such that $\mu - \mu_k \le (\epsilon/2)\mu$. Let $T(X) = X$ if $X \le \tau$ and 0 otherwise. Notice that for X drawn from $\mathcal{D}$, we have $\mathrm{E}[T(X)] \ge (1 - \epsilon/2)\mu$ (by definition of τ). Thus by the law of large numbers we have that there exists an N_3' such that for all $N \ge N_3'$, the following holds.

$$\Pr\left[\sum_{i=1}^{N} T(X_i) \le (1-\epsilon)N\mu\right] \le \epsilon/2. \tag{8}$$

Now set $\epsilon_1 = \sum_{j > \tau} p(j)/2$. Then the probability that X has value at least τ is at least $2\epsilon_1$. Thus, applying the law of large numbers again, we find that there exists an N_3'' s.t. for all $N \ge N_3''$, the following holds:

$$\Pr\left[|\{i | X_i \ge \tau\}| < \epsilon_1 N\right] \le \epsilon/2. \tag{9}$$

Thus, for $N_3 = \max\{N_3', N_3''\}$ and any $N \ge N_3$, we have that with probability at least $1 - \epsilon$ neither of the events mentioned in (8) or (9) occur. In such a case, consider any set S of cardinality at least $(1 - \epsilon_1)N$, and let S' be the set of the $(1 - \epsilon_1)N$ smallest X_i's. We have

$$\begin{aligned}\sum_{i \in S} X_i &\ge \sum_{i \in S'} X_i \\ &\ge \sum_{i=1}^{N} T(X_i) \\ &\ge (1-\epsilon)N\mu.\end{aligned}$$

This proves the claim. ■

C.5 Computing Cesaro Limits in the Transient Case

Lemma C.22 *Let the entries of M and $\vec{\alpha}$ be l-bit rationals describing a transient $(M, \vec{\alpha})$-precess and let π be its stationary probability vector. For every $\epsilon > 0$, there exists $\beta > 0$, with $\log \frac{1}{\beta} = \mathrm{poly}(n, l, \log \frac{1}{\epsilon})$, such that given any vector $\vec{r'}$ of l'-bit rationals satisfying $\|\vec{r'} - \vec{r}\|_\infty \le \beta$, a vector π' satisfying $\|\pi' - \pi\|_\infty \le \epsilon$ can be found in time $\mathrm{poly}(n, l, l', \log \frac{1}{\epsilon})$.*

Proof. Let $\vec{r'}$ be such that $\|\vec{r'} - \vec{r}\|_\infty \le \beta$ (where β will be specified later). We will assume (without loss of generality) that for every i, $r_i' \ge r_i$.

Let D, D_α and X be as in the proof of Lemma 4.13. Define D', D_α' and X' analogously. I.e., D' is the matrix given by

$$D'_{ij} \frac{(1-\alpha_i)M_{ij}}{1 - (1-\alpha_j)\sum_k M_{jk} r_k'},$$

and D' can be described as

$$D' = \begin{bmatrix} D_\alpha' & X' \\ 0 & 0 \end{bmatrix},$$

where D_α' is irreducible. Notice first that $X' = X$, since for any pair i, j s.t. $\alpha_j = 1$, $D_{ij} = D_{ij}' = M_{ij}(1 - \alpha_i)$. Recall our goal is to approximate the maximal left eigenvector π of D, s.t. $\|\pi\|_1 = 1$. Write $\pi = \frac{1}{1+l_B}[\pi_A \pi_B]$, where π_A is a left eigenvector of D_α with $\|\pi_A\|_1 = 1$, $\pi_B = \pi_A X$ and $l_B =$

$\|\pi_B\|_1$. We will show how to compute π'_A, π'_B such that that $\|\pi'_A\|_1 = 1$, $\|\pi'_A - \pi_A\|_\infty \le \epsilon/(n+1)$ and $\|\pi'_B - \pi_B\|_\infty \le \epsilon/(n+1)$. It follows then that if we set $\pi' = \frac{1}{1+\|\pi'_B\|_1}[\pi'_A \pi'_B]$, then

$$\begin{aligned}
\|\pi' - \pi\|_\infty &\le \frac{1}{1+l_B}\max\{\|\pi'_A - \pi_A\|_\infty, \|\pi'_B - \pi_B\|_\infty\} + |l_B - \|\pi'_B\|_1| \\
&\le \frac{\epsilon}{n+1} + |\|\pi'_B - \pi_B\|_1| \\
&\le \epsilon
\end{aligned}$$

as desired.

Further, if π'_A is any vector such that $\|\pi'_A - \pi_A\|_\infty \le \frac{\epsilon}{n(n+1)}$, then a π'_B satisfying $\|\pi'_B - \pi_B\|_\infty \le \epsilon/(n+1)$ can be obtained by setting $\pi'_B = \pi'_A X$. (Notice that $\max_{ij}\{X_{ij}\} \le 1$ and thus $|(\pi'_B)_j - (\pi_B)_j| \le \sum_i X_{ij}|(\pi'_A)_i - (\pi_A)_i| \le n\frac{\epsilon}{n(n+1)}$.)

Thus, below we show how to find π'_A that closely approximates π_A, specifically satisfying $\|\pi'_A - \pi_A\|_\infty \le \epsilon/(n(n+1))$. Notice that this amounts to finding a left eigenvector of the matrix D. We will show how to approximate this using the matrix D'.

We first show that the entries of D' are close to those of D, using the fact that $|r'_k - r_k| \le \beta$. Assume, for notational simplicity, that $r'_k \ge r_k$. Note that

$$\begin{aligned}
D'_{ij} - D_{ij} &= \frac{(1-\alpha_i)M_{ij}}{1-(1-\alpha_j)\sum_k M_{jk}r'_k} - \frac{(1-\alpha_i)M_{ij}}{1-(1-\alpha_j)\sum_k M_{jk}r_k} \\
&= (1-\alpha_i)M_{ij}\frac{(1-\alpha_j)\sum_k M_{jk}(r'_k - r_k)}{(1-(1-\alpha_j)\sum_k M_{jk}r'_k)\,(1-(1-\alpha_j)\sum_k M_{jk}r_k)} \\
&\le \frac{\beta}{(1-(1-\alpha_j)\sum_k M_{jk}r_k)^2}.
\end{aligned}$$

Thus to upper bound this difference, we need an lower bound on the quantity $1-(1-\alpha_j)\sum_k M_{jk}r_k$. If $\alpha_j \ne 0$, then this quantity is at least $\alpha_j \ge 2^{-l}$. Now consider the case where $\alpha_j = 0$. In such a case, for any k, either $\alpha_k = r_k = 1$, or $\alpha_k < 1$ and in such a case, we claim $r_k \le 1 - 2^{-2nl}$. This is true, since the (M,α)-process is irreducible and hence there is a path consisting only of forward moves that goes from k to j, and this path has probability at least 2^{-2nl}, and once we push j onto the history stack, it will never be revoked. Further, by the irreducibility of the $(M,\vec{\alpha})$ process, we have that there must exist a k such that $M_{jk} > 0$ and $r_k \le 1 - 2^{-2nl}$. Using $M_{jk} \ge 2^{-l}$ and substituting, we get that the quantity $1-(1-\alpha_j)\sum_k M_{jk}r_k$ is lower bounded by $2^{-(2n+1)l}$. Thus we conclude that

$$|D'_{ij} - D_{ij}| \le 2^{(4n+2)l}\beta.$$

Next consider the matrix $B = \left(\frac{1}{2}(I + D_\alpha)\right)^n$. Notice that B has a (maximal) eigenvalue of 1, with a left eigenvector π_A. We claim B is positive, with each entry being at least $2^{-(2l+1)n}$. To see this, first note that every non-zero entry of D_α is at least 2^{-2l}. Next consider a sequence $i_0 = i, i_1, i_2, \ldots, i_l = j$ of length at most n satisfying $D_{i_k, i_{k+1}} > 0$. Such a sequence does exist since D_α is irreducible. Further B_{ij} is at least $2^{-n}\prod_k D_{i_k i_{k+1}}$ which is at least $2^{-n(l+1)}$. Thus B is a positive matrix and we are interested in computing its left eigenvector. Lemma C.23 shows how this may be computed given a close approximation to the matrix B.

Next we show that $B' = \left(\frac{1}{2}(I + D'_\alpha)\right)^n$ is a close enough approximation to B. Note that since $\max_{ij}|D_{ij} - D'_{ij}| \le 2^{(4n+2)l}\beta$, we have $\max_{ij}|B'_{ij} - B_{ij}| \le (1 + 2^{(4n+2)l}\beta)^n - 1)$, which may be bounded from above by $(2^n \cdot 2^{(4n+2)l})\beta$ provided $\beta \le 2^{-(4n+2)l}$.

Now let π'_A be any vector satisfying $\|\pi'_A - \pi'_A B'\|_\infty \leq 2^{n+l(4n+2)}\beta$ and $\|\pi'_A\|_1 = 1$. (Such a vector does exist. In particular, π_A satisfies this condition. Further, such a vector can be found by linear programming.) Applying Lemma C.23 to $B^T, (B')^T, \pi_A$ and π'_A with $\gamma = 2^{-n(l+1)}$, $\epsilon = \delta = 2^{n+l(4n+2)}\beta$ yields $\|\pi'_A - \pi_A\|_\infty \leq \sqrt{\beta}2^{O(nl)}$. Thus setting $\beta = \epsilon^2/2^{-\Omega(nl)}$ suffices to get π'_A to be an $\epsilon/(n(n+1))$ close approximation to π_A. This concludes the proof. ■

Lemma C.23 *Let B, C be $n \times n$ matrices and $\hat{x}, \hat{y}$ be n-dimensional vectors satisfying the following conditions:*

1. *For every i, j, $B_{ij} \geq \gamma > 0$, further $\rho(B) = 1$.*
2. *For every i, j, $|C_{ij} - B_{ij}| < \delta$.*
3. *$\|\hat{x}\|_1 = 1$ and $B\hat{x} = \hat{x}$.*
4. *$\|\hat{y}\|_1 = 1$ and $\|C\hat{y} - \hat{y}\|_\infty \leq \epsilon$.*

Then $\|\hat{x} - \hat{y}\|_\infty \leq (1 + \sqrt{n})\sqrt{\frac{2(\epsilon\sqrt{n}+\delta)}{\gamma^3 n}}$.

Proof. We first convert the statement above into one about ℓ_2 norms. Let $x = \frac{\hat{x}}{\|\hat{x}\|_2}$ and $y = \frac{\hat{y}}{\|\hat{y}\|_2}$. Notice that

$$\|Cy - y\|_2 \leq \sqrt{n}\|Cy - y\|_\infty = \frac{\sqrt{n}}{\|\hat{y}\|_2 \|C\hat{y} - \hat{y}\|_\infty} \leq \frac{\sqrt{n}}{\|\hat{y}\|_2}\epsilon \leq \epsilon\sqrt{n}.$$

Thus applying Claim C.24 with $\epsilon' = \epsilon\sqrt{n}$ yields that $\|x - y\|_2 \leq \sqrt{\frac{2(\epsilon\sqrt{n}+\delta)}{\gamma^3 n}}$. Now applying Claim C.25 to the vectors x and y and noticing $\hat{x} = \frac{x}{\|x\|_1}$ and $\hat{y} = \frac{y}{\|y\|_1}$ gives the desired bound. ■

Claim C.24 *Let B, C be $n \times n$ matrices and x, y be n-dimensional vectors satisfying the following conditions:*

1. *For every i, j, $B_{ij} \geq \gamma > 0$; and further $\rho(B) = 1$.*
2. *For every i, j, $|C_{ij} - B_{ij}| < \delta$.*
3. *$\|x\|_2 = 1$ and $Bx = x$.*
4. *$\|y\|_2 = 1$ and $\|Cy - y\|_2 \leq \epsilon'$.*

Then $\|x - y\|_2 \leq \sqrt{\frac{2(\epsilon'+\delta)}{\gamma^3 n}}$.

Proof. Note that B is positive and since x is a non-negative eigenvector, 1 is a maximal eigenvalue of B. This fact is often used below.

Roughly the proof uses standard numerical analysis methods and implicitly goes through the following steps: (1) Argues that the maximal right eigenvector x and the maximal left eigenvector, say z, of the matrix B have a large inner product. (2) Use this to argue that the second eigenvalue of B is small. (3) Resolve the vector y into two components, one parallel to x and the other orthogonal to z, and argue, using the second eignevalue of B, that the component orthogonal to z is small. (4) Argue that y and x must be close, if the component of y orthogonal to z is small. Details below.

1. $\|By - y\|_2 \le (\epsilon' + \delta)$: (This will get rid of C is all future steps.) Let $B = C + \Delta$. Notice that $|\Delta_{ij}| \le \delta$.

$$\begin{aligned} \|By - y\|_2 &= \|Cy + \Delta y - y\|_2 \\ &\le \|Cy - y\|_2 + \|\Delta y\|_2 \\ &\le \epsilon' + \delta\|y\|_2 \\ &= \epsilon' + \delta \end{aligned}$$

2. Let z be the left maximal eigenvector of B, with $\|z\|_2 = 1$. Then $z^T x \ge \gamma^2 n$: We argue this by arguing that every coordinate of z and x is at least γ. Consider x_i, the ith coordinate of x. Let v_i be the ith row of B. Note every coordinate of v_i is at least γ and $x_i = v_i^T x$. Thus $x_i \ge \gamma\|x\|_1 \ge \gamma$. Similarly we can argue $z_i \ge \gamma$. Thus $z^T x \ge \gamma^2 n$ as desired.

3. Let $L = \frac{1}{z^T x} x z^T$, $\beta < \gamma^3 n$, and $B' = B - \beta L$. Then $B' > 0$ and has a maximal eigenvalue equal to $1 - \beta$.

   To see that B' is positive, first notice that every entry of xz^T is at most 1, since $\|x\|_2, \|z\|_2 \le 1$. Next we have that any entry of βL is at most $\frac{\beta}{\gamma^2 n} < \gamma$ since $z^T x$ is at least $\gamma^2 n$. Since every entry of B is at least γ, the claim on the positivity of B' follows.

   Next we notice that x is a non-negative eigenvector of B' with eigenvalue $1 - \beta$, since $B'x = Bx - \frac{\beta}{z^T x}(xz^T)x = x - \frac{\beta}{z^T x}x(z^T x) = x - \frac{\beta}{x} = (1-\beta)x$. This step follows by the Perron-Frobenius theorem (Theorem A.1) which says that $1 - \beta$ is the unique maximal eigenvalue of B'.

4. Let $y = y_x^{||} + y_z^{\perp}$, where $y_x^{||} = \frac{y^T z}{x^T z} x$. Then $\left\|y_z^{\perp}\right\|_2 \le \frac{\epsilon' + \delta}{\gamma^3 n}$.

   Notice that the choice of the vector $y_x^{||}$ makes $y_x^{||}$ parallel to x and $y_z^{\perp}$ orthogonal to z. For the latter, notice

$$(y_z^{\perp})^T z = y^T z - (y_x^{||})^T z = y^T z - \frac{y^T z}{x^t z} x^T z = 0.$$

   We use this fact, or actually the equivalent fact $z^T y_z^{\perp} = 0$ to bound $\left\|y_z^{\perp}\right\|_2$ below. First we note that:

$$\begin{aligned} By - y &= B(y_x^{||} + y_z^{\perp}) - (y_x^{||} + y_z^{\perp}) \\ &= (By_x^{||} - y_x^{||}) + (By_z^{\perp} - y_z^{\perp}) \\ &= \frac{y^T z}{x^T z}(Bx - x) + (By_z^{\perp} - y_z^{\perp}) \\ &= By_z^{\perp} - y_z^{\perp} \\ &= (B' + \beta L)y_z^{\perp} - y_z^{\perp} \\ &= (B'y_z^{\perp} - y_z^{\perp}) + \frac{\beta}{z^T x} x z^T y_z^{\perp} \\ &= (B'y_z^{\perp} - y_z^{\perp}) \end{aligned}$$

   We now use the fact that $\|B'v\|_2 \le \rho(B')\|v\|_2$ for l every vector v, and $\rho(B') = 1 - \beta$ to claim that $\left\|B'y_z^{\perp}\right\|_2 \le (1-\beta)\left\|y_z^{\perp}\right\|_2$. Thus

$$\epsilon' + \delta \ge \|y - By\|_2 = \left\|y_z^{\perp} - B'y_z^{\perp}\right\|_2 \ge \left\|y_z^{\perp}\right\|_2 - \left\|B'y_z^{\perp}\right\|_2 \ge (1 - (1-\beta))\left\|y_z^{\perp}\right\|_2 = \beta\left\|y_z^{\perp}\right\|_2.$$

   This step follows by noticing β can be any real number smaller than $\gamma^3 n$.

5. $\|y-x\|_2 \le \sqrt{\frac{2(\epsilon'+\delta)}{\gamma^3 n}}$:

The crucial observation underlying this step is that the length of the projection of y on the direction orthogonal to x is no larger than vector y above. Note that the projection of y onto the direction orthogonal to x is given by $y-(x^T y)x$. We notice

$$\begin{aligned} y-(x^T y)x &= (y_x^{\|} + y_z^{\perp}) - (x^T(y_x^{\|} + y_z^{\perp}))x \\ &= (y_x^{\|} - (x^T y_x^{\|})x) + (y_z^{\perp} - (x^T y_z^{\perp})x) \\ &= (y_x^{\|} - y_x^{\|}) + (y_z^{\perp} - (x^T y_z^{\perp})x) \\ &= y_z^{\perp} - (x^T y_z^{\perp})x; \end{aligned}$$

Thus

$$\left\| y-(x^T y)x \right\|_2 = \left\| y_z^{\perp} - (x^T y_z^{\perp})x \right\|_2 \le \left\| y_z^{\perp} \right\|_2 \le \frac{\epsilon'+\delta}{\gamma^3 n},$$

where the first inequality uses the fact that the projection of any vector v onto a direction orthogonal to a unit vector u has length less than the length of v.

Applying triangle inequality to the LHS above, we get

$$\|y\|_2 - (x^T y)\|x\|_2. \le \frac{\epsilon'+\delta}{\gamma^3 n}.$$

Using $\|x\|_2 = \|y\|_2 = 1$, we get

$$(x^T y) \ge 1 - \frac{\epsilon'+\delta}{\gamma^3 n}.$$

Now using the fact that $\|y-x\|_2 = \sqrt{2-2(x^T y)}$, we get $\|y-x\|_2 \le \sqrt{\frac{2(\epsilon'+\delta)}{\gamma^3 n}}$.

This concludes the proof of the claim. ■

Claim C.25 *Let $x, y \in \Re^n$ satisfy $\|x\|_2, \|y\|_2 = 1$ and $\|x-y\|_2 \le \delta$. Then $\left\| \frac{x}{\|x\|_1} - \frac{y}{\|y\|_1} \right\|_\infty \le \delta(1+\sqrt{n})$.*

Proof. First we observe $\|x-y\|_1 \le \sqrt{n}\|x-y\|_2 \le \delta\sqrt{n}$. Similarly, $\|x-y\|_\infty \le \|x-y\|_2 \le \delta$. Finally, $\|x\|_1, \|y\|_1 \ge 1$ and $\|y\|_\infty \le 1$. The claim now follows from the following sequence of inequalities.

$$\begin{aligned} \left\| \frac{x}{\|x\|_1} - \frac{y}{\|y\|_1} \right\|_\infty &\le \left\| \frac{x}{\|x\|_1} - \frac{y}{\|x\|_1} \right\|_\infty + \left\| \frac{y}{\|x\|_1} - \frac{y}{\|y\|_1} \right\|_\infty \\ &\le \frac{1}{\|x\|_1}\|x-y\|_\infty + \left| \frac{1}{\|x\|_1} - \frac{1}{\|y\|_1} \right| \|y\|_\infty \\ &\le \delta + \left| \frac{\|x\|_1 - \|y\|_1}{\|x\|_1\|y\|_1} \right| \\ &\le \delta(1+\sqrt{n}). \end{aligned}$$

■