The Annals of Applied Probability
2001, Vol. 11, No. 3, 810–862

RANDOM WALKS WITH "BACK BUTTONS"[1]

BY RONALD FAGIN, ANNA R. KARLIN, JON KLEINBERG, PRABHAKAR RAGHAVAN, SRIDHAR RAJAGOPALAN, RONITT RUBINFELD, MADHU SUDAN[2], AND ANDREW TOMKINS

IBM Almaden Research Center, University of Washington, Cornell University, Verity, Inc., IBM Almaden Research Center, NECI, Massachusetts Institute of Technology and IBM Almaden Research Center

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 by a "forward step." 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 limit distribution, and we give algorithms to compute this distribution. Unlike Markov chains, the limit distribution may depend on the start state.

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, 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 $V = \{1, 2, \dots, n\}$ of states, together with an n-dimensional vector $\boldsymbol{\alpha}$ of *backoff probabilities*. The process evolves as follows: at each time step $t = 0, 1, 2, \dots$, 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 $\operatorname{top}(H)$ denote the top of the stack H. At time $t = 0$ 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: (1) if H_t consists of the singleton X_0 it makes a forward step; (2) otherwise, with probability $\alpha_{\operatorname{top}(H_t)}$ it makes a backward

Received April 1999.

[1]Extended abstract appeared in *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, 484–493. Portland, Oregon, May 2000.

[2]Supported in part by an MIT-NEC Research Initiation Award, a Sloan Foundation Fellowship and NSF Career Award CCR-98-75511.

AMS 2000 *subject classifications.* 60G20, 60J10, 60J22.

Key words and phrases. Denumerable Markov chains, limiting distributions, stochastic processes, worldwide web.

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 equals $\text{top}(H_t)$].

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

Under what conditions do such processes have limit distributions, and how do such processes differ from traditional Markov chains? We focus in this paper on the time-averaged limit distribution, usually called the "Cesaro limit distribution." The *Cesaro limit* of a sequence $a_0, a_1, \ldots$ is $\lim_{t\to\infty}(1/t)\sum_{\tau=0}^{t-1} a_\tau$, if the limit exists. For example, the sequence $0, 1, 0, 1, \ldots$ has Cesaro limit $1/2$. The *Cesaro limit distribution* at state i is $\lim_{t\to\infty}(1/t)\sum_{\tau=0}^{t-1}\Pr[X_\tau = i]$, if the limit exists. By contrast, the *stationary distribution* at state i is $\lim_{t\to\infty}\Pr[X_t = i]$, if the limit exists. Of course, a stationary distribution is always a Cesaro limit distribution. Intuitively, the stationary distribution gives the limiting fraction of time spent in each state, whereas the Cesaro limit distribution gives the average fraction of time spent in each state. We shall sometimes refer simply to either a stationary distribution or a Cesaro limit distribution as a limit distribution.

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, 10, 13] 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 target advertising better on the Web based on a user's browsing patterns; see [3] 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.

Overview of Results. For the remainder of this paper we assume a finite number of states. For simplicity, we assume also that the underlying Markov chain is *irreducible* (i.e., it is possible, with positive probability, to eventually reach each state from every other state) and aperiodic. In particular, M has a stationary distribution and not just a Cesaro limit distribution. Since some backoff probability α_i may equal 1, these assumptions do not guarantee that the backoff process is irreducible (or aperiodic). We are mainly interested in the situation where the backoff process is irreducible. We would like to make the simplifying assumptions that no α_i equals 1 and that the backoff process is irreducible, but we cannot, since later we are forced to deal with cases where these assumptions do not hold.

We now give the reader a preview of some interesting and arguably unexpected phenomena that emerge in such "back-button" random walks. 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 $\boldsymbol{\alpha}$ is identically zero). On the other hand, if the elements of $\boldsymbol{\alpha}$ 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, even if the backoff process is aperiodic. Consider first the case when all entries of $\boldsymbol{\alpha}$ 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 limit distribution provided $\alpha < 1/2$; the history grows unboundedly with time, and the limit 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 limit distribution differs in general from that of M and depends on the start state X_0. Thus, unlike ergodic Markov chains, the limit distribution depends on the start state.

More generally, consider starting the backoff process in a probability distribution over the states of M; then the limit distribution depends on this initial distribution. As the initial distribution varies over the unit simplex, the set of limit distributions forms a simplex. As α converges to 1/2 from above, these simplices converge to a single point, which is the limit 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 limit distribution depend on M and α? Even when all the backoff probabilities are less than 1/2, why should there be a limit 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 limit probabilities. If the process is either ergodic or null then the limit probabilities are computed exactly by solving certain systems of linear inequalities. However, if the process is transient, then the limit probabilities need not be rational numbers, even if all entries of M and α are rational. We show that in this case, the limit 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 limit probability vector.

In Section 2, we establish various definitions and notation. In Section 3, we consider the case where the backoff probabilities are constant (that is, uniform). In Section 4, we consider the general case, where the backoff probabilities can vary. In Section 4.1, we show how it is possible to classify, in polynomial time, the behavior of each irreducible backoff process as transient or ergodic or null. In Section 4.2, we prove that each backoff process always has a Cesaro limit distribution. In Section 4.3, we show how the limit distribution may be computed. In Section 5, we show how it is possible to extend our results to a situation where the backoff probabilities are determined by the edges (that is, for each forward step from state j to state k, there is a probability of revocation that depends on both j and k, rather than depending only on k). In Section 6, we give our conclusions. We also have an Appendix, in which we give some background material, namely, the Perron–Frobenius theorem, Azuma's inequality for martingales, submartingales and supermartingales, the renewal theorem and the law of large numbers. Also in the Appendix, we complete the proof of one of our theorems.

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

For a Markov chain M, backoff vector $\boldsymbol{\alpha}$, and history $\bar{\sigma}$ with $\mathsf{top}(\bar{\sigma}) = j$, define the successor (or next state) $\mathrm{succ}(\bar{\sigma})$ to take on values from $\mathscr{S}$ 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 $\boldsymbol{\alpha}$ and state $i \in \{1, \ldots, n\}$, the $(M, \boldsymbol{\alpha}, i)$-*Markov chain* is the sequence $\langle H_0, H_1, H_2, \ldots \rangle$ taking values from the set $\mathscr{S}$ 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 $X_t = \mathsf{top}(H_t)$ as the $(M, \boldsymbol{\alpha}, i)$-*backoff process*. Several properties of the $(M, \boldsymbol{\alpha}, i)$-backoff process are actually independent of the start state i, and to stress this aspect we will sometimes simply use the term "$(M, \boldsymbol{\alpha})$-backoff process."

Note that the $(M, \boldsymbol{\alpha}, i)$-backoff process does not completely give the $(M, \boldsymbol{\alpha}, i)$-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, even if the backoff process is aperiodic). We shall also study the question of efficiently computing the Cesaro limit distribution.

3. Constant backoff probability. 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. In this case, we refer to the $(M, \boldsymbol{\alpha}, i)$-backoff process as the (M, α, i)-backoff process.

We fix a specific (M, α, 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, α, 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* $\Pr[X_t = j \mid \ell(H_t) = \lambda] = q_\lambda(j)$.

PROOF. Consider a string ω of F's and B's with the property that 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 some $\alpha_i = 0$ or 1). Now consider strings ω and ω' such that $\omega = \omega_1 FB\omega_2$ and $\omega' = \omega_1\omega_2$. Let ω be of length t and ω_1 of length t_1. Notice that

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

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 $\omega = \omega_1 FB\omega_2$ and $\omega' = \omega_1\omega_2$. A sequence ω reduces to a sequence ω'' 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

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

In addition to the sequences $\{X_t\}$ and $\{S_t\}$, consider the sequence $\{Y_t\colon t \geq 0\}$, 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 (except at 0) is α, the probability of moving right (except at 0) is $1 - \alpha$, and the probability of moving right at 0 is 1. 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 $\alpha \le \frac{1}{2}$ and $\alpha > \frac{1}{2}$, and show that in each case there is a Cesaro limit distribution.

The case of $\alpha \le \frac{1}{2}$. Let the stationary probability distribution of the underlying Markov chain M be $\langle \psi_1, \dots, \psi_n \rangle$. By our assumptions about M, this distribution is independent of the start state i. When $\alpha \le \frac{1}{2}$, we show that the (M, α, i)-backoff process converges to $\langle \psi_1, \dots, \psi_n \rangle$. That is, there is a stationary probability distribution, which is independent of the start state i, and this stationary probability distribution equals the stationary probability distribution of the underlying Markov chain.

THEOREM 3.3. *For all states j of the (M, α, i)-backoff process, we have* $\lim_{t\to\infty} \Pr[X_t = j] = \psi_j$.

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

$$\begin{aligned}
&|\Pr[X_t = j] - \psi_j| \\
&\quad = \left| \sum_r q_r(j) \cdot \Pr[Y_t = r] - \psi_j \sum_r \Pr[Y_t = r] \right| \\
&\quad = \left| \sum_r (q_r(j) - \psi_j) \cdot \Pr[Y_t = r] \right| \le \sum_r |q_r(j) - \psi_j| \cdot \Pr[Y_t = r] \\
&\quad = \sum_{r<t_1} |q_r(j) - \psi_j| \cdot \Pr[Y_t = r] + \sum_{r\ge t_1} |q_r(j) - \psi_j| \cdot \Pr[Y_t = r] \\
&\quad \le \sum_{r<t_1} \Pr[Y_t = r] + \sum_{r\ge t_1} \varepsilon/2 \cdot \Pr[Y_t = r] \le \varepsilon/2 + \varepsilon/2 = \varepsilon. \qquad \square
\end{aligned}$$

Although the proof above applies to each $\alpha \le \frac{1}{2}$, 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 $Y_t \le r$; 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, α, 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, α, 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. We now give an example where there is no stationary distribution, even though the backoff process is aperiodic.

EXAMPLE. Assume

$$M = \begin{pmatrix} 0.01 & 0.99 \\ 0.99 & 0.01 \end{pmatrix}, \qquad \alpha = 0.99. \tag{1}$$

Assume that the two states are states 1 and 2, and that the start state is 1. It is easy to see that the backoff process $(M, \alpha, 1)$ can have the initial history $\langle 1 \rangle$ only on even steps. By considering, as before, the corresponding random walk on the natural numbers, with a reflecting barrier at 0, in which the probability of moving left (except at 0) is 0.99, the probability of moving right (except at 0) is 0.01, and the probability of moving right at 0 is 1, we see that on even steps, with high probability the backoff process has the initial history $\langle 1 \rangle$, and hence is in state 1, while on the odd steps, with high probability the backoff process has the history $\langle 1, 2 \rangle$, and hence is in state 2. Since the backoff process is in state 1 with high probability on even steps and is in state 2 with high probability on odd steps, it follows that there is no stationary distribution.

Note that the backoff process is aperiodic: this follows immediately from the fact that there is a self-loop (in fact, both states have a self-loop; that is, it is possible to pass from each state to itself in one step). This is in spite of the fact that there is a periodicity in the histories. Later, we shall study the Markov chain (the "Polish matrix") whose states consist of the attainable histories: the Polish matrix is always periodic.

Now, more generally, suppose that the process starts from an initial distribution over states; we are given a probability vector $z = \langle z_1, \ldots, z_n \rangle$, choose a state j with probability z_j and begin the process from j. As z ranges over all possible probability vectors, what are the possible vectors of limit distributions? Let us again assume a fixed underlying Markov chain M, and denote this set of limit 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.*

PROOF. Let us define $\zeta_r^{(\alpha)}$ to be the value of ζ_r given in the proof of Theorem 3.4 when the backedge probability is α. Define q_t^z to be the probability vector whose jth entry is the probability that the Markov process given by M is in state j after t steps, if the process starts from an initial distribution z of states. Thus, $q_t^z = zM^t$. Note that $q_t(j)$, as defined earlier, is the jth entry of q_t^z when z is the probability distribution with $z_j = 1$ and $z_k = 0$ when $k \neq j$. Define $f_\alpha(z)$ to be the Cesaro limit distribution, when α is the backedge probability. As in the proof of Theorem 3.4, we have $f_\alpha(z) = \sum_r \zeta_r^{(\alpha)} q_r^z$. It is easy to see that f_α is a linear function, which implies that S_α is a simplex.

Let ψ be the stationary probability distribution of the underlying Markov chain M, so that $\psi M = \psi$. We now show that as α converges to $\frac{1}{2}$ from above, the simplices S_α converge to the single vector ψ. We first show that $\psi \in S_\alpha$. Since $\psi M = \psi$, we have $q_t^\psi = \psi$ for every t. It follows easily that $f_\alpha(\psi) = \psi$. Hence, $\psi \in S_\alpha$, as desired.

To show that the S_α's converge to ψ, we show that for each $\varepsilon > 0$, there is α' such that if $\frac{1}{2} < \alpha < \alpha'$, then S_α is in the ball of radius ε about ψ.

We know that $q_t^z = zM^t$ converges to ψ as t goes to infinity for each probability vector z. This convergence is in fact uniform over all probability vectors. That is, given $\varepsilon > 0$, there is T such that for every $t > T$ and for every probability vector z, we have $\|q_t^z - \psi\|_2 < \varepsilon$ (here $\|\cdot\|_2$ is the ℓ_2-norm). Choose k so that $\|q_k^z - \psi\|_2 < \varepsilon/3$ for every probability vector z. Then choose $\alpha' > \frac{1}{2}$ so that for every α with $\frac{1}{2} < \alpha < \alpha'$, we have $\sum_{r<k} \zeta_r^{(\alpha)} < \varepsilon/3$ [it is easy to see that this is possible, by definition of $\zeta_r^{(\alpha)}$]. Then $\|f_\alpha(z) - \psi\|_2 = \|\sum_{r<k} \zeta_r^{(\alpha)}(q_r^z - \psi) + \sum_{r\geq k} \zeta_r^{(\alpha)}(q_r^z - \psi)\|_2 \leq \sum_{r<k} \zeta_r^{(\alpha)} \|q_r^z - \psi\|_2 + \sum_{r\geq k} \zeta_r^{(\alpha)} \|q_r^z - \psi\|_2$. Now $\|q_r^z - \psi\|_2 \leq 2$, since q_r^z and ψ are each probability vectors, and so $\sum_{r<k} \zeta_r^{(\alpha)} \|q_r^z - \psi\|_2 \leq 2\sum_{r<k} \zeta_r^{(\alpha)} < 2\varepsilon/3$. Further, $\|q_r^z - \psi\|_2 < \varepsilon/3$ for $r \geq k$, and so $\sum_{r\geq k} \zeta_r^{(\alpha)} \|q_r^z - \psi\|_2 \leq (\varepsilon/3)\sum_{r\geq k} \zeta_r^{(\alpha)} \leq (\varepsilon/3)\sum_r \zeta_r^{(\alpha)} = \varepsilon/3$. So $\|f_\alpha(z) - \psi\|_2 < \varepsilon$. Therefore, S_α is in the ball of radius ε about ψ, as desired. □

4. Varying backoff probabilities. Recall that the state space $\mathscr{S}$ of the $(M, \boldsymbol{\alpha}, i)$-Markov chain contains all finite attainable histories of the backoff process. Let us refer to the transition probability matrix of the $(M, \boldsymbol{\alpha}, i)$-Markov chain as the *Polish matrix with start 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., [9]). A Markov chain is called *recurrent* if, started in an arbitrary state i, the probability of eventually returning to state i is 1. Otherwise, it is called *transient*. There are two subcases of the recurrent case. If, started in an arbitrary state i, the expected time to return to i is finite, then the Markov chain is called *ergodic*. If, started in an arbitrary state i, the probability of return to state i is 1, but the expected time to return to i is infinite, then the Markov chain is called *null*. Every irreducible Markov chain is either transient, ergodic or null, and for irreducible Markov chains, we can replace every occurrence of "an arbitrary state" by "some state" in these definitions above. Every irreducible Markov chain with a finite state space is ergodic.

As examples, consider a random walk on the natural numbers, 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 $(M, \boldsymbol{\alpha}, i)$ 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 ergodic. The next proposition says that the classification does not depend on the start state and therefore we may refer to the backoff process $(M, \boldsymbol{\alpha})$ as being transient, ergodic or null.

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

PROOF. Let us call a state i *transient* if $(M, \boldsymbol{\alpha}, i)$ 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 $\langle i, \dots, j\rangle$ 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 $\langle i, \dots, j\rangle$. There is now a positive probability that the stack never unwinds back to $\langle i, \dots, j\rangle$ (this follows from the fact that j is transient and that $\alpha_j < 1$). But if the stack never unwinds to $\langle i, \dots, j\rangle$, 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 $(M, \boldsymbol{\alpha}, i)$, the backoff process eventually reaches state j and then makes a forward step. Since the expected time in $(M, \boldsymbol{\alpha}, j)$ to return to the stack $\langle j \rangle$ is infinite, it follows that the expected time in $(M, \boldsymbol{\alpha}, i)$ 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 $(M, \boldsymbol{\alpha}, i)$, after starting in state i. Thus, h_i is finite if i is ergodic, and infinite if i is null. From the start state j in $(M, \boldsymbol{\alpha}, j)$, the expected time to return to the stack $\langle j \rangle$ is

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

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 $M_{j\ell}(1-\alpha_\ell)\alpha_\ell \times (h_\ell + 2)$ represents the situation where the first step is to some state ℓ, 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 $M_{j\ell}(1-\alpha_\ell)^2 \times \alpha_\ell(2h_\ell + 2)$ represents the situation where the first step is to some state ℓ, 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 (2), the expected time to return to the stack $\langle j \rangle$, is finite, so j is ergodic, as desired. □

We shall prove the following theorems.

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

THEOREM 4.3. *Each $(M, \boldsymbol{\alpha}, i)$-backoff process has a Cesaro limit distribution. If the process is irreducible and is either transient or null, then this limit distribution is independent of the start state i. Furthermore, the limit distribution is computable in polynomial time if the process is ergodic or null.*

When the $(M, \boldsymbol{\alpha}, i)$-backoff process is transient, the limit probabilities are not necessarily rational in the entries of M and $\boldsymbol{\alpha}$ (see example in Section 4.3.3) and therefore we cannot hope to compute them exactly. Instead, we give an algorithm for approximating these limit probabilities. Specifically, we show the following.

THEOREM 4.4. *Let $(M, \boldsymbol{\alpha}, i)$ be a transient backoff process on n states, and let all entries of M and $\boldsymbol{\alpha}$ be rationals expressible as ratios of ℓ-bit integers. Then given any error bound $\varepsilon > 0$, a vector π' that ε-approximates the limit distribution π (i.e., satisfies $|\pi'_j - \pi_j| \leq \varepsilon$) can be computed in time polynomial in n, ℓ and $\log \frac{1}{\varepsilon}$.*

The next theorem shows the delicate balance of a null backoff process.

THEOREM 4.5. *Let $(M, \boldsymbol{\alpha})$ be an irreducible, null backoff process.*

(i) *If $(M, \boldsymbol{\alpha})$ is modified by increasing some α_j, but leaving M and all other α_i's the same, then the resulting backoff process is ergodic.*
(ii) *If $(M, \boldsymbol{\alpha})$ is modified by decreasing some α_j, but leaving M and all other α_i's the same, then the resulting backoff process is transient.*

PROOF. The first part is Claim 4.28 below. The second part is demonstrated by comments after Claim 4.28. □

In particular, it follows from Theorem 4.5 that null backoff processes form a set of measure 0.

4.1. *Classifying the backoff process.* In this section we show how it is possible to classify, in polynomial time, the behavior of each irreducible $(M, \boldsymbol{\alpha})$-backoff process as transient or ergodic or null. In Section 3 (where the backoff probability is independent of the state), except for initial histories the expected length of the history either always grows, always shrinks or always stays the same, 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 1 and 2, with $M_{ij} = 1/2$ for every pair i, j, and with $\alpha_1 = 0.99$ and $\alpha_2 = 0.01$. 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 sum of the weights of the states in the stack is expected to grow (resp., shrink, remain constant) if and only if the length of the history is expected to grow (resp., shrink, remain constant) This hope motivates our next few definitions.

DEFINITION 4.6. For a nonnegative vector $\mathbf{w} = \langle w_1, \ldots, w_n \rangle$ and a history $\bar{\sigma} = \langle \sigma_0, \ldots, \sigma_\ell \rangle$ of a backoff process on n states, define the $\mathbf{w}$-potential of $\bar{\sigma}$, denoted $\Phi_{\mathbf{w}}(\bar{\sigma})$, to be $\sum_{i=1}^{\ell} w_{\sigma_i}$ (i.e., the sum of the weights of the states in the history, except the start state).

DEFINITION 4.7. For a nonnegative vector $\mathbf{w} = \langle w_1, \ldots, w_n \rangle$ and a history $\bar{\sigma} = \langle \sigma_0, \ldots, \sigma_\ell \rangle$ of a backoff process on n states, define the $\mathbf{w}$-differential of $\bar{\sigma}$, denoted $\Delta\Phi_{\mathbf{w}}(\bar{\sigma})$, to be $\mathrm{E}[\Phi_{\mathbf{w}}(\mathrm{succ}(\bar{\sigma}))] - \Phi_{\mathbf{w}}(\bar{\sigma})$. (Here E represents the expected value over the distribution given by $\mathrm{succ}(\bar{\sigma})$.)

The following proposition is immediate from the definition.

PROPOSITION 4.8. *If $\bar{\sigma}$ and $\bar{\sigma}'$ are noninitial histories with the same top state j, then*

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

The above proposition motivates the following definition.

DEFINITION 4.9. For a nonnegative vector $\mathbf{w} = \langle w_1, \ldots, w_n \rangle$, a history $\bar{\sigma} = \langle \sigma_0, \ldots, \sigma_\ell \rangle$ of a backoff process on n states and state $j \in \{1, \ldots, n\}$, let $\Delta\Phi_{\mathbf{w},j} = \Delta\Phi_{\mathbf{w}}(\bar{\sigma})$, where $\bar{\sigma}$ is any history with $j = \mathrm{top}(\bar{\sigma})$ and $\ell(\bar{\sigma}) > 0$. Let $\Delta\Phi_{\mathbf{w}}$ denote the vector $\langle \Delta\Phi_{\mathbf{w},1}, \ldots, \Delta\Phi_{\mathbf{w},n} \rangle$.

For intuition, consider the constant α case with weight $w_i = 1$ for each state i. In this case $\Phi_{\mathbf{w}}(\bar{\sigma})$, the $\mathbf{w}$-potential of $\bar{\sigma}$, is precisely $\ell(\bar{\sigma})$, and $\Delta\Phi_{\mathbf{w}}(\bar{\sigma})$, the $\mathbf{w}$-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 varying α case, we would like to associate a positive weight with every state so that (1) the expected change in potential, or the "drift" of the potential, in every step has the same sign independent of the top state (i.e., $\mathbf{w}$ is positive and $\Delta\Phi_{\mathbf{w}}$ is either all positive or all zero or all 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.

EXAMPLES. We now give examples where this approach succeeds in classifying the backoff process as transient, ergodic or null. Let M and $\boldsymbol{\alpha}$ be as follows:

$$M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad \boldsymbol{\alpha} = \langle \tfrac{2}{5}, \tfrac{2}{3} \rangle. \tag{3}$$

Let $\mathbf{w} = (3, 1)$. Then $\Delta\Phi_{\mathbf{w}} = (0, 0)$. Thus, $\mathbf{w}$ is a witness to $(M, \boldsymbol{\alpha})$ being null. (Such results are formally proved later.)

Now let M be as in (3), and let $\boldsymbol{\alpha}' = (\frac{1}{3}, \frac{2}{3})$. Let $\mathbf{w}' = (4, 1)$. Then $\Delta\Phi_{\mathbf{w}'} = (\frac{1}{3}, \frac{1}{6})$. Since $\Delta\Phi_{\mathbf{w}'}$ is all positive, we conclude that $\mathbf{w}'$ is a witness to $(M, \boldsymbol{\alpha}')$ being transient. Note that this is consistent with the second part of Theorem 4.5, since α' is obtained from α by lowering α_1.

Finally, let M be as in (3), and let $\boldsymbol{\alpha}'' = (\frac{2}{5}, \frac{4}{5})$. Let $\mathbf{w}'' = (4, 1)$. (It is a coincidence that $\mathbf{w}' = \mathbf{w}''$.) Then $\Delta\Phi_{\mathbf{w}''} = (-\frac{1}{10}, -\frac{3}{10})$. Since $\Delta\Phi_{\mathbf{w}''}$ is all negative, we conclude that $\mathbf{w}''$ is a witness to $(M, \boldsymbol{\alpha}'')$ being ergodic. Note that this is consistent with the first part of Theorem 4.5, since α'' is obtained from α by raising α_2.

There are easy counterexamples, say, if some $\alpha_i = 1$ and some other $\alpha_j = 0$, that show that it is not possible to insist that the expected change in potential be always positive, or always zero, or always negative, when all weights are positive. Therefore, we relax the requirement of positivity on vectors 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 4.10. We say that an n-dimensional vector $\mathbf{v}$ is admissible for a vector $\boldsymbol{\alpha}$ if $\mathbf{v}$ is nonnegative and $v_i = 0$ only if $\alpha_i = 1$. (We will say simply "admissible" instead of "admissible for $\boldsymbol{\alpha}$" if $\boldsymbol{\alpha}$ is fixed or understood.)

In Section 4.1.1 we prove three very natural lemmas that combine to show the following. Given an irreducible backoff process and an admissible vector $\mathbf{w}$: (1) (Lemma 4.14). If $\Delta\Phi_w$ is admissible then the process is transient. (2) (Lemma 4.19). If $\Delta\Phi_w$ is zero then the process is null. (3) (Lemma 4.17). If $-\Delta\Phi_w$ is admissible then the process is ergodic. Roughly speaking, we show that $\Phi_{\mathbf{w}}(\bar{\sigma})$ 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 $\mathbf{w}$. It does not explain how to find $\mathbf{w}$ or even why the three cases above are exhaustive. In the rest of this section, we show that the cases are indeed exhaustive and give a efficient algorithm to compute $\mathbf{w}$. This part of the argument relies on the surprising properties of an $n \times n$ matrix related to the $(M, \boldsymbol{\alpha})$-process. We now define this matrix, which we call the *Hungarian matrix.*

Let A be the $n \times n$ diagonal matrix with the ith diagonal entry being α_i. Let I be the $n \times n$ identity matrix. If $\alpha_i > 0$ for every i, then the Hungarian matrix for the $(M, \boldsymbol{\alpha})$-process, denoted $H = H_{(M, \boldsymbol{\alpha})}$, is the matrix $(I - A)MA^{-1}$. (Notice that A^{-1} does exist and is the diagonal matrix with ith entry being $1/\alpha_i$.)

The spectral properties of H, and in particular its maximal eigenvalue, denoted $\rho(H)$, play a central role in determining the behavior of the $(M, \boldsymbol{\alpha})$-process. In this section we show how it determines whether the process is ergodic, null or transient. In later sections, we will use it to compute limit probability vectors, for a given $(M, \boldsymbol{\alpha})$-process.

The maximal eigenvalue $\rho(H)$ motivates us to define a quantity $\rho(M, \boldsymbol{\alpha})$ which is essentially equal to $\rho(H)$, in cases where H is defined. Let

$$\rho(M, \boldsymbol{\alpha}) = \sup\{\rho \mid \text{There is an admissible } \mathbf{w} \text{ such that the vector } (I - A)M\mathbf{w} - \rho A\mathbf{w} \text{ is admissible}\}.$$

We first dispense with the case where some $\alpha_i = 0$.

CLAIM 4.11. *If $(M, \boldsymbol{\alpha})$ is irreducible and $\alpha_j = 0$ for some j, then $\rho(M, \boldsymbol{\alpha}) = \infty$.*

REMARK. From the proof it follows that if every entry of M and $\boldsymbol{\alpha}$ is an ℓ-bit rational, then for any $\rho \leq 2^{\ell}$, there exists a nonnegative vector $\mathbf{w}$ with $\|w\|_{\infty} \leq 1$ and $w_i \geq 2^{-\mathrm{poly}(n, \ell)}$ if $w_i \neq 0$ satisfying $(I - A)M\mathbf{w} \geq \rho\mathbf{w}$. This fact will be used in Section 4.3.3.

PROOF. Let $\rho < \infty$ be any constant. We prove the claim by explicitly constructing an admissible vector $\mathbf{w}$ such that $(I - A)M\mathbf{w} - \rho A\mathbf{w}$ is admissible.

Let $M_{\min}$ be the smallest nonzero entry of M, and let $\alpha_{\max}$ be the largest entry of $\boldsymbol{\alpha}$ that is strictly smaller than 1. Let γ be any positive number less than $(1 - \alpha_{\max})M_{\min}/\rho$. Let j be any index such that $\alpha_j = 0$. Let $G_{M, \boldsymbol{\alpha}}$ be the graph on vertex set $\{1, \ldots, n\}$ that has an edge from i to k, if $\alpha_i \neq 1$ and $M_{ik} \neq 0$. [This is the graph with edges giving forward steps of positive probability of the $(M, \boldsymbol{\alpha})$-process.] Let $d(i, k)$ denote the length of the shortest path from i to k in the graph $G_{M, \boldsymbol{\alpha}}$. By the irreducibility of the $(M, \boldsymbol{\alpha})$-process we have that $d(i, j) < n$ for every state i. We now define $\mathbf{w}$ as follows:

$$w_i = \begin{cases} 0, & \text{if } \alpha_i = 1, \\ \gamma^{d(i, j)}, & \text{otherwise.} \end{cases}$$

It is clear by construction that $\gamma > 0$ and thus $\mathbf{w}$ is admissible. Let $\mathbf{v} = (I - A)M\mathbf{w} - \rho A\mathbf{w}$. We argue that $\mathbf{v}$ is admissible componentwise, showing that v_i satisfies the condition of admissibility for every i.

Case 1 ($\alpha_i = 1$). In this case it suffices to show $v_i \geq 0$. This follows from the facts that $\sum_k (1 - \alpha_i)M_{ik}w_k \geq 0$, and $-\rho\alpha_i w_i = 0$ since $w_i = 0$.

Case 2 ($\alpha_i = 0$). (This includes the case $i = j$.). In this case, again we have $-\rho\alpha_i w_i = 0$. Further we have $\sum_k (1 - \alpha_i)M_{ik} = \sum_k M_{ik} = 1$ and thus $v_i = 1$, which also satisfies the condition for admissibility.

Case 3 ($0 < \alpha_i < 1$). In particular, $i \neq j$ and $d(i, j) > 0$. Let k be such that $d(k, j) = d(i, j) - 1$ and there is an edge from i to k in $G_{M,\boldsymbol{\alpha}}$. We know such a state k exists (by definition of shortest paths). We have

$$\begin{aligned} v_i &= \sum_{k'} (1 - \alpha_i) M_{ik'} w_{k'} - \rho \alpha_i w_i \\ &\geq (1 - \alpha_i) M_{ik} w_k - \rho \alpha_i w_i \\ &\geq (1 - \alpha_{\max}) M_{\min}\, w_k - \rho w_i \\ &= (1 - \alpha_{\max}) M_{\min}\, \gamma^{d(k, j)} - \rho \gamma^{d(i, j)} \\ &= \left((1 - \alpha_{\max}) M_{\min} - \rho\gamma\right) \gamma^{d(k, j)} \\ &> 0 \quad \left(\text{since } \gamma < \tfrac{(1-\alpha_{\max}) M_{\min}}{\rho}\right). \end{aligned}$$

Again the condition for admissibility is satisfied. □

The next claim shows that in the remaining cases, $\rho(M, \boldsymbol{\alpha}) = \rho(H)$.

CLAIM 4.12. *Let $(M, \boldsymbol{\alpha})$ be irreducible. If $\alpha_i > 0$ for every i, then $\rho(M, \boldsymbol{\alpha}) = \rho(H)$. Further, there exists an admissible vector $\mathbf{w}$ such that $(I - A)M\mathbf{w} = \rho(M, \boldsymbol{\alpha})A\mathbf{w}$.*

PROOF. Note first that the Hungarian matrix H is nonnegative. Our hope is to apply the Perron–Frobenius theorem to this nonnegative matrix and derive some benefits from this. However, H is not necessarily irreducible, so we can do this yet. So we consider a smaller matrix, $H|_{\boldsymbol{\alpha}}$, which is the restriction of H to rows and columns corresponding to j such that $\alpha_j < 1$. Notice that $H|_{\boldsymbol{\alpha}}$ is irreducible. (This is equivalent to $M|_{\boldsymbol{\alpha}}$ 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 $\mathbf{v}'$ and a (unique) positive real $\rho = \rho(H|_{\boldsymbol{\alpha}})$ such that $H|_{\boldsymbol{\alpha}}\mathbf{v}' = \rho\mathbf{v}'$. In what follows we see that $\rho(M, \boldsymbol{\alpha}) = \rho(H|_{\boldsymbol{\alpha}}) = \rho(H)$.

First we verify that $\rho(H|_{\boldsymbol{\alpha}}) = \rho(H)$. This is easily seen to be true. Note that the rows of H that are omitted from $H|_{\boldsymbol{\alpha}}$ are all 0. Thus a vector $\mathbf{x}$ is a right eigenvector of H if and only if it is obtained from a right eigenvector $\mathbf{x}'$ of $H|_{\boldsymbol{\alpha}}$ by padding with zeroes (in indices j where $\alpha_j = 1$), and this padding preserves eigenvalues. In particular, we get that $\rho(H) = \rho(H|_{\boldsymbol{\alpha}})$ and there is an admissible vector $\mathbf{v}$ (obtained by padding $\mathbf{v}'$) such that $H\mathbf{v} = \rho(H)\mathbf{v}$.

Next we show that $\rho(M, \boldsymbol{\alpha}) \geq \rho(H)$. Consider any $\rho' < \rho(H)$ and let $\mathbf{w} = A^{-1}\mathbf{v}$. Then note that $(I - A)M\mathbf{w} - \rho' A\mathbf{w} = H\mathbf{v} - \rho'\mathbf{v} = (\rho(H) - \rho')\mathbf{v}$ which is admissible. Thus $\rho(M, \boldsymbol{\alpha}) \geq \rho'$ for every $\rho' < \rho(H)$ and thus $\rho(M, \boldsymbol{\alpha}) \geq \rho(H)$.

Finally we show that $\rho(M, \boldsymbol{\alpha}) \leq \rho(H)$. Let $\mathbf{w}$ be an admissible vector and let $\rho > 0$ be such that $(I - A)M\mathbf{w} - \rho A\mathbf{w}$ is admissible. Let $\mathbf{v} = A^{-1}\mathbf{w}$. First note that v_j must be 0 if $\alpha_j = 1$, or else the jth component of the vector $(I - A)M\mathbf{w} - \rho A\mathbf{w}$ is negative. Now let $\mathbf{v}'$ be obtained by restricting $\mathbf{v}$

to coordinates such that $\alpha_j < 1$. Notice now that we have $H|_{\alpha}\mathbf{v}' - \rho\mathbf{v}'$ is a nonnegative vector. From the fact [11], page 17, that

$$\rho(A) = \max_{\mathbf{x}} \left\{ \min_{i|x_i \neq 0} \left\{ \frac{(Ax)_i}{x_i} \right\} \right\}$$

for any irreducible nonnegative matrix A, we conclude that $\rho(H|_{\alpha}) \geq \rho$.

This concludes the proof that $\rho(M, \boldsymbol{\alpha}) = \rho(H)$. The existence of a vector $\mathbf{w}$ satisfying $(I - A)M\mathbf{w} = \rho(H)A\mathbf{w}$ also follows from the argument above. □

LEMMA 4.13. *For every irreducible $(M, \boldsymbol{\alpha})$-backoff process, the following hold*:

(i) *$(M, \boldsymbol{\alpha})$ is ergodic $\Leftrightarrow$ $\rho(M, \boldsymbol{\alpha}) < 1$ $\Leftrightarrow$ there is an admissible $\mathbf{w}$ such that $-\Delta\Phi_{\mathbf{w}}$ is admissible.*

(ii) *$(M, \boldsymbol{\alpha})$ is null $\Leftrightarrow$ $\rho(M, \boldsymbol{\alpha}) = 1$ $\Leftrightarrow$ there is an admissible $\mathbf{w}$ such that $\Delta\Phi_{\mathbf{w}} = \mathbf{0}$.*

(iii) *$(M, \boldsymbol{\alpha})$ is transient $\Leftrightarrow$ $\rho(M, \boldsymbol{\alpha}) > 1$ $\Leftrightarrow$ there is an admissible $\mathbf{w}$ such that $\Delta\Phi_{\mathbf{w}}$ is admissible.*

Furthermore, $\rho(M, \boldsymbol{\alpha})$ and the vector $\mathbf{w}$ are computable in polynomial time.

PROOF. The fact that $\rho(M, \boldsymbol{\alpha})$ is efficiently computable follows from Claims 4.11 and 4.12.

Notice now that $\Delta\Phi_{\mathbf{w}} = (I - A)M\mathbf{w} - A\mathbf{w}$. We start with the case $\rho(M, \boldsymbol{\alpha}) < 1$. Notice that in this case, no $\alpha_i = 0$ (by Claim 4.11) and hence we can apply Claim 4.12 to see that there exists a vector $\mathbf{w}$ such that $(I - A) \times M\mathbf{w} = \rho A\mathbf{w}$. For this vector $\mathbf{w}$, we have $\Delta\Phi_{\mathbf{w}} = (\rho - 1)A\mathbf{w}$. Thus, the vector $-\Delta\Phi_{\mathbf{w}} = (1 - \rho)\mathbf{w}$ is admissible. Applying Lemma 4.17 of Section 4.1.1, we conclude that the $(M, \boldsymbol{\alpha})$-process is ergodic.

Similarly, if $\rho(M, \boldsymbol{\alpha}) = 1$, we have that for the vector $\mathbf{w}$ from Claim 4.12, $\Delta\Phi_{\mathbf{w}} = \mathbf{0}$. Thus, by Lemma 4.19, we find that the $(M, \boldsymbol{\alpha})$-process is null. Finally, if $\rho(M, \boldsymbol{\alpha}) > 1$, then [by the definition of $\rho(M, \boldsymbol{\alpha})$] there exists a vector $\mathbf{w}$ and $\rho' > 1$ such that $(I - A)M\mathbf{w} - \rho' A\mathbf{w}$ is admissible. In particular, this implies that the vector $\Delta\Phi_{\mathbf{w}} = (I - A)M\mathbf{w} - A\mathbf{w}$ is also admissible. Applying Lemma 4.14, we conclude that the $(M, \boldsymbol{\alpha})$-process is transient. □

Theorem 4.2 follows immediately from Lemma 4.13.

4.1.1. *Classification based on drift of the potential.* We now state and prove Lemmas 4.14, 4.17 and 4.19, which relate the drift of the potential to the behavior of the backoff process (i.e., whether they are transient, null or ergodic).

LEMMA 4.14. *For an irreducible $(M, \boldsymbol{\alpha})$-backoff process, if there exists an admissible $\mathbf{w}$ s.t. $\Delta\Phi_{\mathbf{w}}$ is also admissible, then the $(M, \boldsymbol{\alpha})$-backoff process is transient.*

PROOF. We start by showing that the potential $\Phi_{\mathbf{w}}(\text{succ}(\text{succ}(\bar{\sigma})))$ has a strictly larger expectation than the potential $\Phi_{\mathbf{w}}(\bar{\sigma})$. 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 $\{\Phi_{\mathbf{w}}(H_t)\}_t$ 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 process never returns to the initial history, thus ruling out the possibility that it is recurrent (ergodic or null). Details below.

CLAIM 4.15. *There exists $\varepsilon > 0$ such that for all sequences $H_0, \ldots, H_t$ of positive probability in the $(M, \boldsymbol{\alpha}, i)$-Markov chain,*

$$\mathrm{E}\big[\Phi(H_{t+2}) - \Phi(H_t)\big] > \varepsilon.$$

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 back off. 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 $w_j M_{ij}$. Let $\varepsilon_1 = w_j M_{ij}$, and ε_2 be the smallest nonzero entry of $\Delta\Phi_{\mathbf{w}}$. We show that the claim holds for $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$.

Notice first that both the quantities $\mathrm{E}[\Phi(H_{t+1}) - \Phi(H_t)]$ and $\mathrm{E}[\Phi(H_{t+2}) - \Phi(H_{t+1})]$ are nonnegative (since $\Delta\Phi_{\mathbf{w}}$ 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 $\mathrm{E}[\Phi(H_{t+1}) - \Phi(H_t)] = \Delta\Phi_{\mathbf{w},\,\text{top}(H_t)} \geq \varepsilon_2$, since $\Delta\Phi_{\mathbf{w}}$ is admissible.

Case 2 ($\alpha_{\text{top}(H_t)} = 1$ and $\ell(H_t) > 1$). Let $H_t = \langle \sigma_0, \ldots, \sigma_{\ell-1}, \sigma_\ell \rangle$. Note that $H_{t+1} = \langle \sigma_0, \ldots, \sigma_{\ell-1} \rangle$. Further, note that $\alpha_{\text{top}(H_{t+1})} < 1$ (since only the top or bottom of the history stack can be states j with $\alpha_j = 1$). Thus, in this case we have $\mathrm{E}[\Phi(H_{t+2}) - \Phi(H_{t+1})] \geq \varepsilon_2$ (again using the admissibility of $\Delta\Phi_{\mathbf{w}}$).

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

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

PROOF. Since the potential at the initial history 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 $\varepsilon t/2$. Further notice that the sequence $\Phi_{\mathbf{w}}(H_0), \Phi_{\mathbf{w}}(H_2), \Phi_{\mathbf{w}}(H_4), \ldots,$ form a submartingale, and that the change in $\Phi_{\mathbf{w}}(H_t)$ is absolutely bounded: $|\Phi_{\mathbf{w}}(H_{t+2}) - \Phi_{\mathbf{w}}(H_t)| \leq 2 \cdot \max_{i\in\{1,\ldots,n\}}\{w_i\}$. Therefore, we can apply a standard tail inequality (Corollary A.5) to show that there exist constants $c < \infty$, $\lambda < 1$ such that

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

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 $(M, \boldsymbol{\alpha}, i)$-backoff 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 $(M, \boldsymbol{\alpha}, i)$-backoff process returns to the initial history after time T is bounded away from 1, ruling out the possibility that it is recurrent.

LEMMA 4.17. *For an irreducible $(M, \boldsymbol{\alpha})$-backoff process, if there exists an admissible $\mathbf{w}$ s.t. $-\Delta\Phi_{\mathbf{w}}$ is also admissible, then the $(M, \boldsymbol{\alpha})$-backoff process is ergodic.*

PROOF. First notice that we can modify the vector $\mathbf{w}$ so that it is positive and $\Delta\Phi_{\mathbf{w}}$ is negative, as follows. Let ε be the smallest nonzero entry of $-\Delta\Phi_{\mathbf{w}}$. For every j such that $\alpha_j = 1$, set $w'_j = w_j + \varepsilon/2$. The corresponding difference vector, $\Delta\Phi_{\mathbf{w}'}$, is at most $\varepsilon/2$ larger than $\Delta\Phi_{\mathbf{w}}$ in any coordinate, and thus entries that were already negative in $\Delta\Phi_{\mathbf{w}}$ remain negative in $\Delta\Phi_{\mathbf{w}'}$. On the other hand, for any j such that $\Delta\Phi_{\mathbf{w},j}$ was 0 (implying $\alpha_j = 1$), the value of $\Delta\Phi_{\mathbf{w}',j}$ is $-w'_j = -\varepsilon/2$. Thus all the zero entries are now negative.

Henceforth we assume, without loss of generality, that $\mathbf{w}$ is positive and $\Delta\Phi_{\mathbf{w}}$ is negative. Let $w_{\min}$ denote the smallest entry of $-\Delta\Phi_{\mathbf{w}}$ and $w_{\max}$ denote the largest entry of $\mathbf{w}$. At this stage the expected $\mathbf{w}$-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 history sequence $H_0, \ldots, H_t, \ldots$ of the $(M, \boldsymbol{\alpha}, i)$-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 $\psi(t) = \psi_{\mathbf{w}}^{H_0,\ldots,H_t,\ldots}(t)$ to be

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

By construction, the extended potential of a sequence is expected to go down by at 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 supermartingale and the change in one step is absolutely bounded. Thus, by applying a martingale tail inequality (Corollary A.6), we see that for any $\varepsilon > 0$, with probability tending to 1, the extended potential after t steps is at most $-(1-\varepsilon)w_{\min} \cdot t$. [More formally, for every $\varepsilon, \delta > 0$, 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 $-(1-\varepsilon)w_{\min}\cdot t$, is at most δ.] Since the $\Phi_{\mathbf{w}}$ part of the extended potential is always nonnegative, and each time the sequence reaches the initial history the extended potential is reduced by at most $(w_{\max}+w_{\min})$, this implies that a sequence with extended potential $-(1-\varepsilon)w_{\min}\cdot t$ must include at least $((1-\varepsilon)w_{\min}/(w_{\min}+w_{\max}))\cdot t$ initial histories.

Assume for contradiction that the $(M,\boldsymbol{\alpha})$-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. By the law of large numbers (Proposition A.9), we find that for every $\delta > 0$ and every c, there exists an integer N such that with probability at least $1-\delta$, the first N visits to the initial history take more than cN steps. Setting $\delta = \frac{1}{2}$ and $c = 2\cdot(w_{\min}+w_{\max})/((1-\varepsilon)w_{\min})$ and $t = cN$, we see from the previous paragraph that with probability tending to 1, after t steps there are at least $2N$ initial histories. But we just showed that with probability at least $\frac{1}{2}$, the first N visits to the initial history take more than t steps. This is a contradiction. We conclude that the $(M,\boldsymbol{\alpha})$-backoff process is ergodic. □

Before going on to characterize null processes, we prove a simple proposition that we will need in the next lemma. Define the *revocation probability* r_j to be the probability that a given forward move to state j is eventually popped off the history stack (for a more formal definition, see Definition 4.36).

PROPOSITION 4.18. *If an irreducible* $(M,\boldsymbol{\alpha})$*-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 process returns to the initial history with probability 1, making it recurrent. □

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

LEMMA 4.19. *For an irreducible* $(M,\boldsymbol{\alpha})$*-backoff process, if there exists an admissible* $\mathbf{w}$ *s.t.* $\Delta\Phi_{\mathbf{w}} = \mathbf{0}$ *then the* $(M,\boldsymbol{\alpha})$*-backoff process is null.*

PROOF. We first define an extended potential as in the proof of Lemma 4.17, but we will be a bit more careful. Let $\tau = \mathrm{E}[\Phi_{\mathbf{w}}(H_1)-\Phi_{\mathbf{w}}(H_0)]$ be the expected increase in potential from the initial history. (Note $\tau > 0$.)

As before, for a history sequence $H_0,\ldots,H_t,\ldots$ of the $(M,\boldsymbol{\alpha},i)$-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

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

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 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 $(M, \boldsymbol{\alpha})$-backoff process fits in neither category.

CLAIM 4.20. *If the irreducible $(M, \boldsymbol{\alpha})$-backoff process is transient, then there exist constants $\varepsilon > 0$ and b such that for every time t, it is the case that*

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

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 step. 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 $(M, \boldsymbol{\alpha}, i)$-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 nonzero entry of $\mathbf{w}$. Then the expected value of $\Phi_{\mathbf{w}}(H_t)$ 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 start 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 4.21. *If the irreducible $(M, \boldsymbol{\alpha})$-backoff process is ergodic, then there exist constants $\gamma > 0$ and b such that for all t,*

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

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 (Proposition A.9), we have that there exists t_0 such that for all $t \geq t_0$ the probability that the number of visits to the initial history 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 $\mathrm{E}[\Phi_{\mathbf{w}}(H_t)]$. Let $f(t)$ denote the smallest nonnegative index such that the history $H_{t-f(t)}$ is an initial history. Notice then that $\mathrm{E}[\Phi_{\mathbf{w}}(H_t)]$ 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 $\varepsilon > 0$, there exists a constant a such that $F(t) \leq \varepsilon t + a$. Set a such that $\sum_{i>a} ip(i) \leq \frac{\varepsilon}{2}T$. The base cases of the induction are with $t \leq a$ and these easily satisfy the hypothesis, since $F(t) \leq t \leq a \leq \varepsilon t + a$. For $t > a$, we get

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

By selecting ε sufficiently small (so that the overall coefficient of t is negative), the claim follows. □

4.2. *Existence of Cesaro limit distributions.* In this section we prove that the (M, α, i)-backoff process always has a Cesaro limit distribution. The proofs are different for each case (ergodic, null and transient), and so we divide the discussion based on the case. In the transient case, we prove even more (the existence of a stationary distribution, not just a Cesaro limit distribution, when the backoff process is aperiodic). As we showed earlier, there need not be a stationary distribution in the ergodic case, even when the backoff process is aperiodic. It is an open problem as to whether there is always a stationary distribution in the aperiodic null case (we conjecture that there is).

4.2.1. *Ergodic case.* The simplest argument is for the ergodic case.

THEOREM 4.22. *If the (M, α, i)-backoff process is ergodic, then it has a Cesaro limit distribution.*

PROOF. 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 states (stacks) in the Polish matrix with top state i. □

4.2.2. *Transient case.* Next, we consider the transient case (where the Polish matrix is transient). The crucial notion underlying the analysis of this case is that of "irrevocability." 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 limit 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 argument is formalized below, to obtain a proof of the existence of a Cesaro limit distribution.

THEOREM 4.23. *If the (M, α, i)-backoff process is transient, then it has a Cesaro limit distribution, which is independent of the start state i. If it is aperiodic, then it has 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 $\text{top}(\bar{\sigma})$ has $\alpha_{\text{top}(\bar{\sigma})} \neq 1$, there is a positive probability, starting in $\bar{\sigma}$, that the top state $\text{top}(\bar{\sigma})$ is never popped off of the stack. It is clear that this probability depends only on the top state $\text{top}(\bar{\sigma})$ 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 Polish 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 $\alpha_i \neq 1$. It is not hard to see this assumption is without loss of generality, since the backoff process will reach an irrevocable state j (which necessarily has $\alpha_j \neq 1$) with probability 1.

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 (if the backoff process is aperiodic) and (most important) 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 $\sigma_2, \ldots,$ and then to σ_ℓ, with each of the states $\sigma_1, \sigma_2, \ldots, \sigma_\ell$ being nonirrevocable (they can be nonirrevocable, since it is possible to back up from σ_ℓ to $\sigma_{\ell-1} \cdots$ 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 if the backoff process 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, or does not take place at all after time t_0). It is sufficient to show that the probability of E_k is $O(\theta^k)$ for some constant $\theta < 1$. 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 size of the stack (state) in the Polish matrix is one bigger than it was at time t_0 infinitely often.

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 size of the stack in the Polish matrix at time $t_0 + k'$ is one bigger than it was at time t_0.

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), there are positive constants c_1, c_2 such that the probability that the size of the stack in the Polish matrix has grown by only one after k' steps is at most $c_1 e^{-c_2 k'}$. Therefore, the probability that there is some $k' \geq k$ such that that the size of the stack in the Polish matrix has grown by only one after k' steps is at most $\sum_{k' \geq k} c_1 e^{-c_2 k'} = c_1 e^{-c_2 k} \sum_{m=0}^{\infty} e^{-c_2 m} = c_1(1/(1 - e^{-c_2}))e^{-c_2 k}$. Let $\theta = e^{-c_2}$. So the probability of E_k is $O(\theta^k)$, as desired.

We have shown that the expected time between irrevocable states is finite. So starting in state $\langle \sigma_0 \rangle$ of the Turkish matrix, 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 and ergodic. So it has a Cesaro limit distribution. This gives us a Cesaro limit distribution in the backoff process, where the probability of state m is the sum of the probabilities of the stacks in the Turkish matrix with top state m. Since the Turkish matrix is the same, independent of the start state i, this probability does not depend on the start state. (As mentioned earlier, there is a subtlety if the start state i has $\alpha_i = 1$. It is not hard to see that this independence of the Cesaro limit probabilities on the start state holds even then.) If the backoff process is aperiodic, then the Turkish matrix has a stationary distribution, and hence so does the backoff process. □

4.2.3. *Null case.* Finally, we consider the null case. In this case our proof is based on a surprising property of a recurrent (ergodic or null) (M, α, i)-backoff process: its steady state distribution turns out to be independent of α_i (the backoff probability of the start state). We exploit this property (which will be proved implicitly in Lemma 4.25 below) as follows. 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 is 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 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.

THEOREM 4.24. *If the (M, α, i)-backoff process is null, then it has a Cesaro limit distribution, which is independent of the start state i.*

As before, we can assume without loss of generality that the (M, α, i)-backoff process is irreducible, since we can easily restrict our attention to an irreducible "component."

The theorem follows from Lemma 4.25 below, which asserts that the limit distribution exists and equals the limit distribution of a related ergodic process and is independent of the start state i.

LEMMA 4.25. *Let (M, α) be null. Let j be any state of M such that $\alpha_j < 1$. Let α' be the vector given by $\alpha'_j = 1$ and $\alpha'_{j'} = \alpha_{j'}$ otherwise. Then the (M, α', j)-backoff process is ergodic and hence has a Cesaro limit distribution. Let i be any state of M. Then the (M, α, i)-backoff process has a Cesaro limit distribution which is the same as the Cesaro limit distribution of the (M, α', j)-backoff process.*

PROOF. The first part of Lemma 4.25, claiming that (M, α') is ergodic, follows from the first part of Theorem 4.5, and is proven in Claim 4.28. We now move to the more difficult part. It is convenient for us to use the term *walk*, which refers to the sequence of states visited (along with the information about the auxiliary sequence that tells whether each move was a forward or backward step and the history sequence). For this part, we consider a walk W of length t of the (M, α, i)-backoff 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, α, 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 $S_{t_1} = \text{F}$ (i.e., W makes a forward step at time t_1) and $X_{t_1} = j$, we define a partition of W into two walks W' and W'' as follows.

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 $\langle 0, \ldots, t_1, t_2+1, \ldots, t \rangle$ of time steps (and the associated sequence of states visited and auxiliary sequences). They give a new $(M, \boldsymbol{\alpha}, i)$-walk W' that has positive probability. On the other hand, the sequence $\langle t_1, t_1+1, \ldots, t_2 \rangle$ of time steps defines a walk W'' of an $(M, \boldsymbol{\alpha}, j)$-backoff 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-divisions of W. Specifically, W is a j-decomposition of itself. Further, if (a) $W_0, \ldots, W_\ell$ is a j-decomposition of W', (b) $W_{\ell+1}, \ldots, W_k$ is a j-decomposition of W'', and (c) (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 nontrivial 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 j-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 $j \neq i$):

(i) Initial skeleton. This is a skeleton that has $\langle i \rangle$ as its initial history. Note that there is exactly one such skeleton (unless $i = j$, in which case we say that there is no initial skeleton). We denote the initial skeleton by W_0.

(ii) Closed skeletons. These are the skeletons with $\langle j \rangle$ as their initial and final history.

(iii) Open skeletons. These are the skeletons with $\langle j \rangle$ as their initial history, but not their final history.

Our strategy for analyzing the frequency of the occurrence 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 $(M, \boldsymbol{\alpha}', j)$-backoff process. But the $(M, \boldsymbol{\alpha}', j)$-backoff process is ergodic, and thus the expected time to return to the initial history in such walks is finite. With a large number of closed j-skeletons, the frequency of occurrence of j' converges (to its frequency in $(M, \boldsymbol{\alpha}', j)$-backoff processes).

Consider the following.

Simulation of W.

1. Pick an (infinite) walk W'_0 from the $(M, \boldsymbol{\alpha}', i)$-backoff process.
2. Pick a sequence $W'_1, W'_2, \ldots,$ of walks as follows: for each k, the walk W'_k starts at $\langle j \rangle$ and walks according to $(M, \boldsymbol{\alpha}', j)$ 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 point.
   (c) Truncate W to its first t steps and output it. Further, let W_i denote the truncation of W_i' so that it includes only the initial portion of W_i' that is used in W.

The following proposition is easy to verify.

PROPOSITION 4.26. *W generated as above has exactly the same distribution as that of the $(M, \boldsymbol{\alpha}, i)$-backoff process. Further $W_0, \dots, W_N$ give the j-skeletal decomposition of W.*

Let W' denote a random walk obtained by starting at $\langle j \rangle$, walking according to $(M, \boldsymbol{\alpha}', j)$ and stopping the first time we reach the initial history. Since the $(M, \boldsymbol{\alpha}', j)$-backoff 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'. By Theorem A.8 $\mu_{j'}/\mu = \pi'_{j'}$, where π' denotes the Cesaro limit distribution of the $(M, \boldsymbol{\alpha}', j)$-backoff 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' ($k \in \{1, \dots, N\}$) 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 now enumerate a number of bad events, argue that each one of them has low probability of occurrence and then argue that if none of them happens, then for N sufficiently large,

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

1. N is too small. In Claim 4.29 we show that this event has low probability. Specifically, there exists $\delta > 0$ such that for every $\varepsilon > 0$ there exists t_0 such that for all $t \geq t_0$, the probability that N is less than δt is at most ε.
2. W_0 is too long. Claim 4.30 shows that for every $\varepsilon > 0$, 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 4.32, we prove that for every $\varepsilon_0 > 0$, there exists t_2 such that if $t \geq t_2$, then the probability that the number of open skeletons is more than $\varepsilon_0 t$ is at most ε_0.

4. $\sum_{k=1}^{N} b_k$ is too large. By the law of large numbers (Proposition A.9), we have that for every $\varepsilon > 0$, there exists N_1 such that for all $N \geq N_1$, the probability that $\sum_{k=1}^{N} b'_k \geq (1 + \varepsilon)\mu N$ is at most ε. Using the fact that $b_k \leq b'_k$, we obtain the same upper bound on $\sum_{k=1}^{N} b_k$ as well.
5. $\sum_{k=1}^{N} a_k$ is too large. As above, we have that we have that for every $\varepsilon > 0$, there exists N_2 such that for all $N \geq N_2$, the probability that $\sum_{k=1}^{N} a_k \geq (1 + \varepsilon)\mu_{j'} N$ is at most ε.
6. (Informally) $\sum_{k=1}^{N} b_k$ is too small. The first formal event considered here is that for some large subset $S \subseteq \{1, \dots, N\}$, 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 4.34 can be used to show that for every $\varepsilon > 0$, there exists $\varepsilon_1 > 0$ and $N_3 > 0$, such that for all $N \geq N_3$ the probability that there exists a subset $S \subseteq \{1, \dots, N\}$ of cardinality at least $(1 - \varepsilon_1)N$ such that $\sum_{k \in S} b'_k \leq (1 - \varepsilon)\mu N$ 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 the negation of item (3), we get to the informal claim here.
7. $\sum_{k=1}^{N} a_k$ is too small. Obtained as above. Specifically, for every $\varepsilon > 0$, there exists $\varepsilon_2 > 0$ and $N_4 > 0$, such that for all $N \geq N_4$ the probability that there exists a subset $S \subseteq \{1, \dots, N\}$ of cardinality at least $(1 - \varepsilon_2)N$ such that $\sum_{k \in S} a'_k \leq (1 - \varepsilon)\mu N$ 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 $\varepsilon = \min\{\beta/7, \beta/(2 + 1/(\mu\delta)), \beta/(2 + 1/(\mu_{j'}\delta) + \beta)\}$. Let ε_1 and ε_2 be as given in items (6) and (7) above and let $\varepsilon_0 = \min\{\varepsilon, \varepsilon_1\delta, \varepsilon_2\delta\}$. 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 take $t \geq \max\{t_0, t_1, t_2, \frac{1}{\delta}N_1, \frac{1}{\delta}N_2, \frac{1}{\delta}N_3, \frac{1}{\delta}N_4\}$. Then since t is large enough, we have that for any of items (1), (2) or (3), the probability that 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 $1 - 7\varepsilon \geq 1 - \beta$. We now reason that if none of these events happen then $\sum_{k=0}^{N} a_k / \sum_{k=0}^{N} b_k$ is between $(1-\beta)\pi'_{j'}$ and $(1+\beta)\pi'_{j'}$. We show the lower bound; the proof of the upper bound is similar. We first give an upper bound for $\sum_{k=0}^{N} b_k$ by the negations of items (2) and (4). By the negation of item (2), $b_0 \leq \varepsilon t \leq \frac{\varepsilon}{\delta}N$ [where the second inequality uses the negation of item (1).] By the negation of item (4), $\sum_{k=1}^{N} b_k \leq (1 + \varepsilon)\mu N$ and thus we have

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

Next we give a lower bound on $\sum_{k=0}^{N} a_k$. Here we use the negation of item (3) to conclude that the number of closed skeletons is at least $N - \varepsilon_0 t \geq N - (\varepsilon_0/\delta)N \geq (1 - \varepsilon_2)N$. Let S denote the set of indices k of closed skeletons

W_k. Thus, we have

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

Putting the above together, we get

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

as desired. [The final inequality above uses $\pi'_{j'} = \mu_{j'}/\mu$ and $\varepsilon \leq \beta/(2+1/(\mu\delta))$.] The upper bound follows similarly, using the inequality $\varepsilon \leq \beta/(2+1/(\mu_{j'}\delta)+\beta)$. This concludes the proof of the lemma, modulo Claims 4.28–4.34. □

For the following claims, let H denote the Hungarian matrix corresponding to the $(M, \boldsymbol{\alpha})$-backoff process and let H' denote the Hungarian matrix corresponding to the $(M, \boldsymbol{\alpha}')$-backoff process. For a nonnegative matrix A, let $\rho(A)$ denote its maximal eigenvalue. For $n \times n$ matrices A and B, say $A < B$ if $A_{ik} \leq B_{ik}$ for every i, k and there exists i, k such that $A_{ik} < B_{ik}$. Claim 4.28 will use the following simple claim.

CLAIM 4.27. *If A and B are $n \times n$ irreducible nonnegative 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 positive integer k, we have $\rho((I+A)^k) < \rho((I+B)^k)$, 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, we have $C_{ij} < D_{ij}$. (By contrast, we know only that the strict inequality $A_{ij} < B_{ij}$ holds for *some* pair i, j.) Notice that the (i, j) 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 weight of a walk is the product of the weight of the edges it takes, and where 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 $I+B$. By assumption, there are ℓ, m so that $A_{lm} < B_{lm}$. 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 $P_1 \circ (\ell, m) \circ P_2$ 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

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

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

$$\begin{aligned}
\rho(D) &= \max_{\mathbf{y}} \min_{i \in \{1,\dots,n\}} \left\{ \frac{(D\mathbf{y})_i}{(\mathbf{y})_i} \right\} \\
&\geq \min_{i \in \{1,\dots,n\}} \left\{ \frac{(D\mathbf{x})_i}{(\mathbf{x})_i} \right\} \\
&> \min_{i \in \{1,\dots,n\}} \left\{ \frac{(C\mathbf{x})_i}{(\mathbf{x})_i} \right\} \quad (\text{since } D_{ij} > C_{ij} \text{ and } \mathbf{x} \neq 0) \\
&= \rho(C) \quad (\text{by our choice of } \mathbf{x}).
\end{aligned}$$

□

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

CLAIM 4.28. *Let $(M, \boldsymbol{\alpha})$ be irreducible and null. Let j be a state such that $\alpha_j < 1$. Assume that $\alpha'_j > \alpha_j$ and $\alpha'_{j'} = \alpha_{j'}$ if $j' \neq j$. Then $(M, \boldsymbol{\alpha}')$ is ergodic (though it may not be irreducible).*

PROOF. We first focus on the case $\alpha'_j < 1$. In this case, we observe that $(M, \boldsymbol{\alpha}')$ is also irreducible. For this part, we use Lemma 4.13 and Claim 4.12 to rephrase this question in terms of the maximal eigenvalues of the corresponding Hungarian matrices H [for $(M, \boldsymbol{\alpha})$] and H' [for $(M, \boldsymbol{\alpha}')$]. In particular, we have $\rho(H) = 1$ and we need $\rho(H') < \rho(H) = 1$.

Note that for every k, ℓ, we have

$$\begin{aligned}
H'_{kl} &= (1 - \alpha'_k) M_{kl} \alpha'^{-1}_{\ell} \\
&\leq (1 - \alpha'_k) M_{kl} \alpha^{-1}_{\ell} \\
&\leq (1 - \alpha_k) M_{kl} \alpha^{-1}_{\ell} \\
&= H_{kl}.
\end{aligned}$$

Further, the first inequality is strict if $\ell = j$ and $M_{kj} \neq 0$ (and such a k does exist, by the irreducibility of M). Using Claim 4.27 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 $(M, \boldsymbol{\alpha}'')$-backoff process, where $\alpha_j < \alpha''_j < 1$ (and $\alpha''_{j'} = \alpha_{j'}$ for other j'), is ergodic. Thus it suffices to prove that $(M, \boldsymbol{\alpha}')$ is ergodic, given that $(M, \boldsymbol{\alpha}'')$ is ergodic. However, since we may not have irreducibility, we need to argue this individually for every $(M, \boldsymbol{\alpha}', i)$-backoff process. Now the expected return time of the $(M, \boldsymbol{\alpha}'', i)$-backoff process (to its initial history) is finite. But it is not hard to see that the expected return time of the $(M, \boldsymbol{\alpha}', i)$-backoff process (to its initial history) is bounded above by the expected return time of the $(M, \boldsymbol{\alpha}'', i)$-backoff process, since it can only cost additional steps to not pop j immediately off the history stack whenever it appears. So the $(M, \boldsymbol{\alpha}', i)$-backoff process is ergodic, as desired. □

Claim 4.28 gives the first part of Theorem 4.5. The proof of the second part is similar, provided α_j is not lowered to $\alpha'_j = 0$ (in which case the Hungarian matrix would not be defined). However, if $\alpha'_j = 0$, then the resulting backoff process is certainly transient.

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

CLAIM 4.29. *There exists $\delta > 0$, such that for every $\varepsilon > 0$ there exists 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 onto 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. 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 (Proposition A.9), we get that there exists t_0 such that 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 $\delta = \rho/(2n)$. □

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

CLAIM 4.30. *For every $\varepsilon > 0$, there exists a time t_1 such that for all times $t > t_1$,*

$$\Pr[\textit{length of } W_0 > \varepsilon t] < \varepsilon.$$

PROOF. We prove the claim in two steps. First we note that in a walk of length t, with high probability, the (null) $(M, \boldsymbol{\alpha}, i)$-backoff 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 the following.

SUBCLAIM 1. For every $\varepsilon' > 0$, there exists a time t'_1 such that for all $t > t'_1$, the probability that an $(M, \boldsymbol{\alpha}, i)$-walk of length t returns to the initial history at least $\varepsilon' t$ times is at most ε'.

The next subclaim follows from the law of large numbers (Proposition A.9).

SUBCLAIM 2. Let T be the expected return time to the initial history in the $(M, \boldsymbol{\alpha}', i)$-backoff process. [Note that $T < \infty$, since the $(M, \boldsymbol{\alpha}', i)$-backoff process is ergodic.] Then for every ε'', there exists N_0 such that if $N \geq N_0$ and $N' \leq N$, then the probability that N' returns to the initial history take more than $2NT$ steps is at most ε''.

From the two subclaims, we get the claim as follows. Set $\varepsilon'' = \varepsilon/2$ and $\varepsilon' = \min\{\varepsilon/2, \varepsilon/(2T)\}$. Now let N_0 and T be as in Subclaim 2 and let $t_0 = \max\{t'_1, 2N_0T/\varepsilon\}$. Given $t \geq t_0$, let $N = (\varepsilon t)/(2T)$. Notice $N \geq N_0$. Applying Subclaim 1, we get that the probability that the number of returns to the initial history in the $(M, \boldsymbol{\alpha}, i)$-backoff process is at least N [$= (\varepsilon t)/(2T) \geq \varepsilon' t$] is at most $\varepsilon' \leq \varepsilon/2$. Now applying Subclaim 2, we get that the probability of N returns to the initial history taking more that $2NT = \varepsilon t$ steps in the $(M, \boldsymbol{\alpha}', i)$-backoff process is at most $\varepsilon'' = \varepsilon/2$. So with probability at least $1-\varepsilon$, neither of the bad events in the subclaims occurs, which means that there are less than N returns to the initial history in the initial skeleton, and even N returns would take time at most εt steps. So with probability at least $1 - \varepsilon$, the length of the initial skeleton is at most εt. This proves the claim. □

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

CLAIM 4.31. *If $(M, \boldsymbol{\alpha}, i)$ is null and $\mathbf{w}$ is a weight vector as guaranteed to exist by Lemma* 4.13, *then the* $\mathbf{w}$*-potential $\Phi_{\mathbf{w}}(H_t)$ is expected to grow as $o(t)$.*

PROOF. Recall that the extended potential used in Lemma 4.17 is expected to be 0 after t steps. Further, by Subclaim 1 of Claim 4.30, the number of returns to the initial history is less than $\varepsilon' t$, with probability all but ε'. Thus the expected number of returns to the initial history is at most $2\varepsilon' t$. Hence, the expected value of $\phi_{\mathbf{w}}(H_t)$ is also at most $2\varepsilon' t$. □

CLAIM 4.32. *For every $\varepsilon > 0$, there exists 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 at time t is at most ε.*

PROOF. Consider the event E that the history H_t contains more than εt occurrences 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 $\mathbf{w}$ be the weight vector as shown to exist in Lemma 4.13, and let $\phi_{\mathbf{w}}(H_t)$ be the potential of the history H_t. Notice that if E occurs, then the potential $\phi_{\mathbf{w}}(H_t)$ is at least $w_j \varepsilon t$. Since E happens with probability at least ε, the expected potential $\mathrm{E}[\phi(H_t)]$ is at least $\varepsilon^2 w_j t$, and so is growing linearly in t. But this contradicts the previous claim. □

We now use our machinery to prove a lemma that we need to prove Theorem 4.35.

LEMMA 4.33. *In a null backoff process, for every $\varepsilon > 0$, there exists t_2 such that for all $t \geq t_2$, the probability that more than εt of the forward edges into state ℓ are unrevoked at time t is at most ε.*

PROOF. If $\alpha_\ell < 1$, then if we take $j = \ell$ every unrevoked edge into state ℓ corresponds, in the machinery we have just developed, to a different open skeleton. The result then follows from Claim 4.32. If $\alpha_\ell = 1$, then every forward edge into state ℓ is immediately followed by a revocation, so the result again follows. □

Our final claim to complete the proof of Lemma 4.25, and hence of Theorem 4.24, is a technical one. In this claim, $[N] = \{1, \dots, N\}$.

CLAIM 4.34. *For every distribution $\mathcal{D}$ on nonnegative integers with finite expectation μ and every $\varepsilon > 0$, there exists $\varepsilon_1 > 0$ and $N_3 > 0$ such that for all $N \geq N_3$, if $X_1, \dots, X_N$ are N samples drawn i.i.d. from $\mathcal{D}$, then*

$$\Pr\left[\sum_{i\in S} X_i \geq (1-\varepsilon)\mu N \text{ for every } S \subseteq [N] \text{ with } |S| \geq (1-\varepsilon_1)N\right] \geq 1-\varepsilon.$$

PROOF. We will find ε_1 and pick τ such that with high probability the $(\varepsilon_1 N)$th largest element of $X_1, \dots, X_N$ 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 $\mu_k = \sum_{j\leq k} jp(j)$. Notice that the μ_k's converge to μ. Let τ be such that $\mu - \mu_\tau \leq (\varepsilon/2)\mu$. Let $T(X) = X$ if $X \leq \tau$ and 0 otherwise. Notice that for X drawn from $\mathcal{D}$, we have $\mathrm{E}[T(X)] \geq (1-\varepsilon/2)\mu$ (by definition of τ). Thus by the law of large numbers (Proposition A.9), there exists N_3' such that for all $N \geq N_3'$, the following holds:

$$\Pr\left[\sum_{i=1}^{N} T(X_i) \leq (1-\varepsilon)N\mu\right] \leq \varepsilon/2. \tag{4}$$

Now set $\varepsilon_1 = \sum_{j>\tau} p(j)/2$. Then the probability that X has value at least τ is at least $2\varepsilon_1$. Thus, applying the law of large numbers (Proposition A.9) again, we find that there exists N_3'' such that for all $N \geq N_3''$, the following holds:

$$\Pr\big[\big|\{i | X_i \geq \tau\}\big| < \varepsilon_1 N\big] \leq \varepsilon/2. \tag{5}$$

Thus, for $N_3 = \max\{N_3', N_3''\}$ and any $N \geq N_3$, we have that with probability at least $1-\varepsilon$ neither of the events mentioned in (4) or (5) occur. In such a case, consider any set S of cardinality at least $(1-\varepsilon_1)N$, and let S' be the set of the $(1-\varepsilon_1)N$ smallest X_i's. We have

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

This proves the claim. □

4.3. *Computation of limit distributions.* We now show how the limit distribution may be computed. We can assume without loss of generality that the backoff process is irreducible, since we can easily restrict our attention to an irreducible "component." Again, we branch into three cases.

4.3.1. *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, α)-backoff process. The next theorem gives an important application of the Hungarian matrix.

THEOREM 4.35. *The limit 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 limit probability distribution can be found by solving a linear system.*

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

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

The only controversial condition is the third one, that $\pi_{i \to j} = \pi_{i \leftarrow j}$. The fact that $\pi_{i \leftarrow j}$ exists and equals $\pi_{i \to j}$ follows easily from Lemma 4.33. Manipulating the above conditions shows that π satisfies $\pi = \pi H$.

We now consider uniqueness. Assume first that $\alpha_i < 1$ for every i. Then H is irreducible and nonnegative and thus by the Perron–Frobenius theorem (Theorem A.1), it follows easily that π is the unique solution to the linear system. If some $\alpha_i = 1$, we argue by focusing on the matrix $H|_\alpha$, which is irreducible (as in Section 4.1, $H|_\alpha$ is the principal submatrix of H containing only rows and columns corresponding to i such that $\alpha_i < 1$). Renumber the states of M so that the α_i's are nondecreasing. Then the Hungarian matrix looks as follows:

$$H = \begin{pmatrix} H|_\alpha & X \\ 0 & 0 \end{pmatrix},$$

where $H|_\alpha$ is nonnegative and irreducible and X is arbitrary. Write $\pi = (\pi_A \pi_B)$, where π_B has the same number of elements as the number of α_i's that are 1. Then the linear system we have to solve is

$$(\pi_A \pi_B) = (\pi_A \pi_B) \begin{pmatrix} H|_\alpha & X \\ 0 & 0 \end{pmatrix}.$$

This system can be solved by finding $\pi_A = \pi_A H|_\alpha$ and then setting $\pi_B = \pi_A X$. Now π_B is uniquely determined by π_A. Furthermore, π_A is uniquely determined by the Perron–Frobenius theorem (Theorem A.1). This concludes the proof of the theorem. □

4.3.2. *Ergodic case.* In this case also the limit 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 back at the initial history, it first flips a coin that comes up B with probability α_i and F otherwise. If the coin comes up B, the end of an epoch is declared.

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 noninitial 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. From Theorem A.8 we see that the Cesaro limit probability distribution vector $\pi^{(i)}$, for an $(M, \boldsymbol{\alpha}, i)$-backoff process, is given by $\pi_j^{(i)} = N_{ij}/T_i$, provided T_i is finite. This gives us a way to compute the Cesaro limit distribution. The key equations that allow us to compute the N_{ij} and T_i are

$$T_i = 1 + \sum_k M_{ik}[\alpha_k \cdot 1 + (1-\alpha_k)(T_k + 1)] + (1-\alpha_i)T_i, \tag{6}$$

$$N_{ij} = \delta_{ij} + \sum_k M_{ik}[\alpha_k \cdot \delta_{jk} + (1-\alpha_k)(N_{kj} + \delta_{jk})] + (1-\alpha_i)N_{ij}, \tag{7}$$

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.]

We claim that the first set (6) of linear equations completely specify T. We argue this as follows. First we may rearrange terms in the equation and use the fact that $\sum_k M_{ik} = 1$ to simplify (6) to

$$\alpha_i T_i = 2 + \sum_k (1-\alpha_k) M_{ik} T_k.$$

Dividing both sides by α_i (we know that no $\alpha_i = 0$ in the ergodic case), moving all terms involving T_k to the left, and using the fact that the Hungarian matrix H is given by $H_{ik} = ((1-\alpha_k)/\alpha_i)M_{ik}$, we get

$$T_i - \sum_k H_{ik} T_k = \frac{2}{\alpha_i}.$$

Letting $\mathbf{T} = \langle T_1, \ldots, T_n \rangle$ and $\mathbf{b} = \langle 2/\alpha_1, \ldots, 2/\alpha_n \rangle$, we get $(I - H)\mathbf{T} = \mathbf{b}$. Since the maximal eigenvalue of H is less than 1, we know that $I - H$ has an inverse (and is given by $I + H + H^2 + \cdots$) and thus $\mathbf{T}$ is given by $(I - H)^{-1}\mathbf{b}$.

Similarly, if we let $\mathbf{N}_j = \langle N_{1j}, \ldots, N_{nj} \rangle$ and $\mathbf{b}_j = \langle (\delta_{1j} + M_{1j})/\alpha_1, \ldots, (\delta_{nj} + M_{nj})/\alpha_n \rangle$, then (7) simplifies to yield $\mathbf{N}_j = (I - H)^{-1}\mathbf{b}_j$.

Thus $\mathbf{T}$ and the $\mathbf{N}_j$'s can be computed using the above linear equations. Using now the formula $\pi_j^{(i)} = N_{ij}/T_i$, we can also compute the stationary probability vectors.

4.3.3. *Transient case.* We now prove Theorem 4.4.

We now give a formal definition of the revocation probability, which was defined informally earlier.

DEFINITION 4.36. For a state j, define the revocation probability as follows. Pick any noninitial history $\bar{\sigma} = \langle \sigma_0, \ldots, \sigma_\ell \rangle$ with $\text{top}(\bar{\sigma}) = j$. The *revocation probability* r_j is the probability that the $(M, \boldsymbol{\alpha}, i)$-Markov chain starting at state $\bar{\sigma}$ reaches the state $\bar{\sigma}' = \langle \sigma_0, \ldots, \sigma_{\ell-1} \rangle$. (Notice that this probability is independent of i, ℓ, and $\sigma_0, \ldots, \sigma_{\ell-1}$; thus, the quantity is well-defined.)

Note that r_i is the probability that an epoch starting at i, as in Section 4.3.2, ends in finite time. Let $\mathbf{r}$ denote the vector of revocation probabilities. The following lemma shows how to compute the limit probabilities π given $\mathbf{r}$. Further, it shows how to compute a close approximation to π, given a sufficiently close approximation to $\mathbf{r}$.

LEMMA 4.37. *The limit probabilities satisfy* $\pi = \pi(I - A)MR$, *where* R *is a diagonal matrix with* $R_{ii} = 1/(1 - (1 - \alpha_i)\sum_k r_k M_{ik})$. *Further, there exists a unique solution to the this system subject to the condition* $\sum_i \pi_i = 1$.

REMARKS. If $\alpha_i = 0$ for every i, then $r_i = 0$ for every i, and so we recover the familiar condition for Markov chains that $\pi = \pi M$. Although we are considering the transient case here, note that if we formally take $r_i = 1$, which occurs in the null case, then we in fact recover the equation we found in the null case, namely $\pi = \pi(I - A)MA^{-1}$.

PROOF. The first part of the lemma is obtained as in Theorem 4.35. Let $\pi_{i \to j}$ denote the limit probability of a forward step from i to j, and let $\pi_{i \leftarrow j}$ denote the limit probability of a backward step from j to i. Then the following conditions hold:

$$\pi_{i \leftarrow j} = r_j \pi_{i \to j}, \tag{8}$$

$$\pi_{i \to j} = \pi_i(1 - \alpha_i)M_{ij}, \tag{9}$$

$$\pi_i = \sum_j \pi_{j \to i} + \sum_j \pi_{i \leftarrow j}. \tag{10}$$

Equation (8) is shown as follows. Note that is is obvious that $\pi_{i \to j}$ exists and equals $\pi_i(1 - \alpha_i)M_{ij}$. Let $\pi^+_{i \leftarrow j}$ denote the limiting probability of moves *from* i to j that are eventually revoked. Note that this limit does exist and equals $r_j \pi_{i \to j}$. Next, using the ergodicity of the Turkish matrix (from the proof of Theorem 4.23) we deduce that $\pi_{i\leftarrow j}$ also exists (it is the sum of certain steady-state transitions of the Turkish matrix). Finally, we note that the total number of unrevoked forward moves at time t that eventually do get revoked is given by the length ℓ of the state $\langle\sigma_0, \ldots, \sigma_\ell\rangle$ in the Turkish matrix. By the ergodicity of the Turkish matrix, this length is $o(t)$ with probability $1 - o(1)$. Thus, we get $\pi_{i\leftarrow j} = \pi^{+}_{i\leftarrow j}$, and this yields Equation (8).

Using (8) to eliminate all occurrences of variables of the form $\pi_{i\leftarrow j}$ and then using (9) to eliminate all occurrences of $\pi_{i\rightarrow j}$, (10) becomes

$$\pi_i = \sum_j \pi_j(1-\alpha_j)M_{ji} + \pi_i(1-\alpha_i)\sum_j r_j M_{ij}. \tag{11}$$

Thus if we let D be the matrix with

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

then π satisfies $\pi = \pi D$. As in the proof of Theorem 4.35 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:

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

where D_α is nonnegative and irreducible. Thus $\pi = [\pi_A \pi_B]$ must satisfy $\pi_A = \pi_A D_\alpha$ and $\pi_B = \pi_A X$. Now π_A is 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 4.38. *Let the entries of M and $\boldsymbol{\alpha}$ be ℓ-bit rationals describing a transient $(M, \boldsymbol{\alpha}, i)$-backoff process and let π be its limit probability vector. For every $\varepsilon > 0$, there exists $\beta > 0$, with* $\log \frac{1}{\beta} = \text{poly}(n, \ell, \log \frac{1}{\varepsilon})$, *such that given any vector $\mathbf{r}'$ of ℓ'-bit rationals satisfying $\|\mathbf{r}' - \mathbf{r}\|_\infty \le \beta$, a vector π' satisfying $\|\pi' - \pi\|_\infty \le \varepsilon$ can be found in time* $\text{poly}(n, \ell, \ell', \log \frac{1}{\varepsilon})$.

REMARK. By truncating $\mathbf{r}'$ 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 4.38 to the Appendix **B**.

In the next lemma, we address the issue of how the revocation probabilities may be determined. We show that they form a solution to a quadratic program, in fact a semidefinite program. [Recall that a real symmetric matrix A is positive semidefinite if all of its eigenvalues are nonnegative. A semidefinite program is an optimization problem with a linear objective function whose constraints are of the form "$A[\mathbf{x}]$ is positive semidefinite," where $A[\mathbf{x}]$ denotes a symmetric matrix whose entries are themselves linear forms in the variables $x_1, \ldots, x_n$. Semidefinite programs are a special case of convex programs, but more general than linear programs. They can be approximately solved efficiently using the famed ellipsoid algorithm (see [4] for more details).]

LEMMA 4.39. *The revocation probabilities r_i are the optimum solution to the following system:*

$$\begin{aligned} &\min \sum_i x_i \\ &\textit{such that } x_i \geq \alpha_i + (1-\alpha_i) x_i \textstyle\sum_j M_{ij} x_j, \\ &\qquad x_i \leq 1, \\ &\qquad x_i \geq 0. \end{aligned} \tag{12}$$

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

$$\begin{aligned} &\min \sum_i x_i \\ &\textit{such that } q_i = 1 - (1-\alpha_i), \textstyle\sum_j M_{ij} x_j, \\ &\qquad x_i \leq 1, \\ &\qquad x_i \geq 0, \\ &\qquad q_i \geq 0, \\ &\qquad D_i \textit{ positive semidefinite, where } D_i = \begin{pmatrix} x_i & \sqrt{\alpha_i} \\ \sqrt{\alpha_i} & q_i \end{pmatrix}. \end{aligned} \tag{13}$$

PROOF. We start by considering the following iterative system and proving that it converges to the optimum of (12).

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

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

By induction, we note that $x_i^{(t)} \leq x_i^{(t+1)} \leq 1$. The first inequality holds, since

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

The second inequality follows similarly. Hence, since $\langle x_i^{(t)} \rangle_t$ is a nondecreasing sequence in the interval [0, 1], it must have a limit. Let x_i^* denote this limit.

We claim that the x_i^* give the (unique) optimum to (12). 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 (12). To prove that it is the optimum, we claim that if $a_1, \ldots, a_n$ are a feasible solution to (12), then we have $a_i \geq x_i^{(t)}$ and thus $a_i \geq x_i^*$. We prove this claim by induction. Assume $a_i \geq x_i^{(t)}$, for every i. Then

$$a_i \geq \alpha_i + (1-\alpha_i) a_i \sum_j M_{ij} a_j$$

$$\geq \alpha_i + (1 - \alpha_i)x_i^{(t)} \sum_j M_{ij} x_j^{(t)}$$

$$= x_i^{(t+1)}.$$

This concludes the proof that the x_i^* give the unique optimum to (12).

Next we show that the revocation probability r_i equals 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 to i is revoked at the first step with probability α_i, or there is a 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 eventually revoked.] Thus the r_i's form a feasible solution, and so $r_i \geq x_i^*$. To prove that $r_i \leq x_i^*$, let us define $r_i^{(t)}$ to be the probability that a forward step 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 $r_i^{(t+1)} \leq x_i^{(t+1)}$, as desired. Thus we conclude that $x_i^* = r_i$. This finishes the proof of the first part of the lemma.

For the second part, note that the condition that D_i be positive semidefinite is equivalent to the condition that $x_i q_i \geq \alpha_i$. Substituting $q_i = 1 - (1 - \alpha_i) \sum_j M_{ij} x_j$ turns this into the constraint $x_i - (1 - \alpha_i) x_i \sum_j M_{ij} x_j \geq \alpha_i$, and thus establishing the (syntactic) equivalence of (12) and (13). □

Using Lemmas 4.37 and 4.39 above, we can derive exact expressions for the revocation probabilities and limit probabilities of any given backoff process. The following example illustrates this. It also shows that the limit probabilities are not necessarily rational, even when the entries of M and $\boldsymbol{\alpha}$ are rational.

EXAMPLE. The following example shows that the limit probabilities may be irrational even when all the entries of M and $\boldsymbol{\alpha}$ are rational. Let M and $\boldsymbol{\alpha}$ be as follows:

$$M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \qquad \boldsymbol{\alpha} = \langle \tfrac{1}{2}, \tfrac{1}{3} \rangle.$$

Using Lemma 4.39, we can now show that the revocation probabilities are roots of cubic equations. Specifically, r_1 is the unique real root of the equation $-16+30x-13x^2+2x^3=0$ and r_2 is the unique real root of the equation $-9+21x-14x^2+8x^3=0$. Both quantities are irrational and given approximately by $r_1 \approx 0.7477$ and $r_2 \approx 0.5775$. Applying Lemma 4.37 to this, we find that the limit probabilities of the (M,α)-process are π_1 and π_2, where π_1 is the unique real root of the equation

$$-1024+3936x-3180x^2+997x^3=0$$

and π_2 is the unique real root of the equation

$$-729+567x+189x^2+997x^3=0.$$

It may be verified that the cubic equations above are irreducible over the rationals, and thus π_1 and π_2 are irrational and given approximately by $\pi_1 \approx 0.3467$ and $\pi_2 \approx 0.6532$.

In the next lemma we show how to efficiently approximate the vector of revocation probabilities. The proof assumes the reader is familiar with standard terminology used in semidefinite programming and in particular the notion of a separation oracle and its use in the ellipsoid algorithm (see [4] for more details).

LEMMA 4.40. *If the entries of M and α are given by ℓ-bit rationals, then an ε-approximation to the vector of revocation probabilities can be found in time* $\mathrm{poly}(n,\ell,\log\frac{1}{\varepsilon})$.

PROOF. We solve the convex program given by (12) approximately using the ellipsoid algorithm [4]. Recall that the ellipsoid algorithm can solve a convex programming problem given (i) a separation oracle describing the convex space, (ii) a point $\mathbf{x}$ inside the convex space, (iii) radii ε and R such that the ball of radius ε around $\mathbf{x}$ 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 $\log\frac{R}{\varepsilon}$.

The fact that (12) describes a convex program follows from the fact that it is equivalent to the semidefinite program (13). Further, a separation oracle can also be obtained due to this equivalence. In what follows we will describe a vector $\mathbf{x}$ that is feasible, and an $\varepsilon \geq 2^{-\mathrm{poly}(n,\ell)}$ such that every point y satisfying $\|x-y\|_\infty \leq \varepsilon$ 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 lemma.

Recall, from Lemma 4.13 of Section 4.1, that since (M,α) is transient, there exists $\rho > 1$ and a vector $\mathbf{w}$ such that $(I-A)M\mathbf{w} \geq \rho A\mathbf{w}$. Let $w_{\max} = \max_i\{w_i\}$ and $w_{\min} = \min_{i|w_i\neq 0}\{w_i\}$. Notice further that we can choose ρ and $\mathbf{w}$ such that $\rho \geq 1+2^{-\mathrm{poly}(n,\ell)}$ and $w_{\max}=1$ and $w_{\min} \geq 2^{-\mathrm{poly}(n,\ell)}$. [In case $\rho(M,\alpha)=\infty$, this follows by picking, say $\rho=2$, and using the remark after Claim 4.11. In case $\rho(M,\alpha)<\infty$ we use Claim 4.12 and set

$\rho = \rho(H)$ and $\mathbf{w} = A^{-1}\mathbf{v}$, where $\mathbf{v}$ is a right eigenvector of H. Since $\rho > 1$ is an eigenvalue of a matrix whose entries are ℓ-bit rationals and since $\mathbf{w}$ is a multiple of the eigenvector, the claims about the magnitude of ρ and $w_{\min}$ follow.]

Before describing the vector $\mathbf{x}$ 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 (12) only for the remaining choices of indices i. So henceforth we assume $0 < \alpha_i < 1$ and in particular the fact that $\alpha_i \geq 2^{-\ell}$.

Let $\delta = (\rho - 1)/(2\rho)$. Note $\delta > 2^{-\mathrm{poly}(n,\ell)}$. Let $\varepsilon = 2^{-(\ell+3)} w_{\min}((\rho - 1)/\rho)^2$. We will set $z_i = 1 - \delta w_i$ and first show that $z_i - \alpha_i - (1 - \alpha_i) z_i \sum_j M_{ij} z_j$ is at least 2ε. Consider

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

Now consider any vector $\mathbf{y}$ such that $z_i - 2\varepsilon \leq y_i \leq z_i$. We claim that $\mathbf{y}$ is feasible. First, $y_i \leq 1$ since $y_i \leq z_i = 1 - \delta w_i \leq 1$. We now show that $y_i \geq 0$. First, $z_i \geq 0$ since $w_i \leq 1$ and $\delta < 1$. Since, as we showed above, $z_i - \alpha_i - (1 - \alpha_i) z_i \sum_j M_{ij} z_j \geq 2\varepsilon$, it follows that $y_i \geq z_i - 2\varepsilon \geq \alpha_i + (1 - \alpha_i) z_i \sum_j M_{ij} z_j \geq 0$. Finally,

$$\begin{aligned}
y_i - \alpha_i - (1-\alpha_i) y_i \sum_j M_{ij} y_j &\geq z_i - 2\varepsilon - \alpha_i - (1-\alpha_i) y_i \sum_j M_{ij} y_j \\
&\geq z_i - 2\varepsilon - \alpha_i - (1-\alpha_i) z_i \sum_j M_{ij} z_j \\
&\geq 0 \quad \text{(using the claim about the } z_i\text{'s).}
\end{aligned}$$

Thus setting $x_i = z_i - \varepsilon$, we note that every vector $\mathbf{y}$ satisfying $x_i - \varepsilon \leq y_i \leq x_i + \varepsilon$ is feasible. This concludes the proof. □

PROOF OF THEOREM 4.4. Given M, $\boldsymbol{\alpha}$ and ε, let β be as given by Lemma 4.38. We first compute a β-approximation to the vector of revocation probabilities in time $\text{poly}(n, \ell, \log \frac{1}{\beta}) = \text{poly}(n, \ell, \log \frac{1}{\varepsilon})$ using Lemma 4.40. The output is a vector $\mathbf{r}'$ of $\ell' = \text{poly}(n, \ell, \log \frac{1}{\varepsilon})$-bit rationals. Applying Lemma 4.38 to M, $\boldsymbol{\alpha}$, $\mathbf{r}$ and ε, we obtain an ε-approximation to the limit probability vector π in time $\text{poly}(n, \ell, \ell', \log \frac{1}{\varepsilon}) = \text{poly}(n, \ell, \log \frac{1}{\varepsilon})$. □

5. Allowing backoff probabilities on edges. In this paper, we have considered the backoff probability to be determined by the current state. What if we were to allow the backoff probabilities to be a function not just of the current state, but of the state from which the current state was entered by a forward step? Thus, in this situation, each edge (j', j) that corresponds to a forward step from j' to j has a probability of being revoked that depends not just on j, but on j' also. We refer to this new, more general process as an *edge-based backoff process*, and our original backoff process as a *node-based backoff process*. We now define edge-based backoff processes a little more precisely.

As with node-based backoff processes, for an edge-based backoff process we are given a Markov matrix M, indexed by the set $\mathscr{S}$ of states. The difference is that for node-based backoff processes, we are given a vector $\boldsymbol{\alpha}$ of backoff probabilities α_i for each state i; however, for edge-based backoff processes, we are given a vector $\boldsymbol{\kappa}$ of backoff probabilities κ_{ij} for each pair i, j of states.

For a history $\bar{\sigma} = \langle \sigma_0, \ldots, \sigma_{\ell-1}, \sigma_\ell \rangle$, define next-to-top$(\bar{\sigma})$ to be $\sigma_{\ell-1}$.

Given the Markov chain M and backoff vector $\boldsymbol{\kappa}$, and history $\bar{\sigma}$ with next-to-top$(\bar{\sigma}) = j'$ and top$(\bar{\sigma}) = j$, define the successor (or next state) succ$(\bar{\sigma})$ to take on values from $\mathscr{S}$ with the following distribution:

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

We denote by $(M, \boldsymbol{\kappa}, i)$ the edge-based backoff process, with start state i.

We now show how to convert our results about node-based backoff processes into results for edge-based backoff processes. Assume we are given the edge-based backoff process $(M, \boldsymbol{\kappa}, i)$. Let $\mathscr{S}'$ be the set of all ordered pairs (j, k) of states of $\mathscr{S}$ such that $M_{jk} > 0$. Define a new matrix M', indexed by $\mathscr{S}'$, such that $M'_{(j,k)(k,\ell)} = M_{k\ell}$, and $M'_{(j,k)(\ell,m)} = 0$ if $\ell \neq k$. It is easy to verify that M' is a Markov chain, that M' is irreducible if M is and that M' is aperiodic if M is. Define $\boldsymbol{\alpha}$ over $\mathscr{S}'$ by taking $\alpha_{(j,k)} = \kappa_{jk}$. We correspond to the edge-based backoff process $(M, \boldsymbol{\kappa}, i)$ the node-based backoff process $(M, \boldsymbol{\alpha}, (m, i))$, where m is an arbitrary state in $\mathscr{S}$ such that $M_{mi} > 0$. This correspondence allows us to carry over results about node-based backoff processes into results about edge-based backoff processes. For example, we have the following result.

THEOREM 5.1. *Every edge-based backoff process has a Cesaro limit distribution.*

PROOF. Let (M, κ, i) be an edge-based backoff process. Let $(M', \alpha, (m, i))$ be the corresponding node-based backoff process. We have shown that every node-based backoff process has a Cesaro limit distribution. This gives a Cesaro limit distribution for the edge-based backoff process, where the limit probability of state j in the edge-based backoff process is the sum of the limit probabilities of all states (k, j) in the corresponding node-based backoff process. □

Similarly, the Cesaro limit distribution for the edge-based backoff process is efficiently computable, just as it is for node-based backoff processes.

6. Conclusions. We have introduced backoff processes, which are generalizations of Markov chains where it is possible, with a certain probability, to back up to the previous state that was entered by a forward step. Backoff processes are intended to capture a feature of browsing on the World Wide Web, namely, the use of the back button, that Markov chains do not. We show that backoff processes have certain properties that are similar to those of Markov chains, along with some interesting differences. Our main focus is on limiting distributions, which we prove always exist and can be computed efficiently.

We view this research as only a first step. First, we believe that backoff processes are a natural extension of Markov chains that deserve further study. Second, we feel that further generalizations should be considered and investigated. We gave one simple example of such a generalization in Section 5. More powerful generalizations should be considered, including studies as to how well various stochastic models actually capture browsing, along with a mathematical analysis of such models.

APPENDIX

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., [7], page 508). *Let A be an irreducible, nonnegative square matrix. Then*:

(i) *There exists* $\mathbf{v}$, *with all components positive, and* $\lambda_0 > 0$ *such that* $A\mathbf{v} = \lambda_0\mathbf{v}$.
(ii) $\lambda_0 = \sup_{\mathbf{x}}\{\min_{i|x_i \neq 0}\{(Ax)_i/x_i\}\}$.
(iii) *If* $\lambda \neq \lambda_0$ *is any other eigenvalue of* A, *then* $|\lambda| < \lambda_0$.
(iv) *Each* $\mathbf{w}$ *such that* $A\mathbf{w} = \lambda_0\mathbf{w}$ *is a constant multiple of* $\mathbf{v}$.
(v) *Each nonnegative eigenvector of* A *is a constant multiple of* $\mathbf{v}$.

A.2. *Martingale tail inequalities.* We begin by reviewing the basic definitions.

DEFINITION A.2. We now define a martingale, supermartingale and submartingale.

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

(ii) A sequence $X_0, X_1, \ldots$ of random variables is said to be a *supermartingale* if $E[X_i | X_0, \ldots, X_{i-1}] \le X_{i-1}$ for all $i > 0$.

(iii) A sequence $X_0, X_1, \ldots$ of random variables is said to be a *submartingale* if $E[X_i | X_0, \ldots, X_{i-1}] \ge X_{i-1}$ for all $i > 0$.

THEOREM A.3 (Azuma's inequality; see, e.g., [12] page 92). *Let $X_0, X_1, \ldots$ be a martingale such that for each k*

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

where c_k may depend on k. Then for each $t \ge 0$ and each $\lambda > 0$,

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

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

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

Then for each $t \ge 0$ and each $\lambda > 0$,

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

COROLLARY A.5. *Let $X_0, X_1, \ldots$ be a submartingale such that*

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

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

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

Then for each $t \ge 0$ and each $\lambda \ge 0$,

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

COROLLARY A.6. *Let $X_0, X_1, \ldots$ be a supermartingale such that*

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

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

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

Then for all $t \ge 0$,

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

A.3. *Renewal theory.*

DEFINITION A.7. A *renewal process* $\{N(t),\ t \geq 0\}$ is a nonnegative integer-valued stochastic process that counts the number of 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., [8] page 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 epoch [period between ith and $(i+1)$st event], 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. *Law of large numbers.* We shall make use various times of the following (weak form of the) law of large numbers.

PROPOSITION A.9. *Let $p\colon \mathcal{Z}^+ \to [0,1]$ 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*

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

B. Stability of computations in the transient case. In this section we show that the linear system used to find the stationary probability vector (given the vector of revocation probabilities) in Section 4.3.3 is stable. Thus it can be solved even if some of the entries of the system are only known approximately. This proof relies on a general theorem (Theorem B.1), due to Gurvits [5], about the stability of the maximal eigenvector of a positive matrix. For completeness, a proof of this theorem is also included in this section.

RESTATEMENT OF LEMMA 4.38. *Let the entries of M and $\boldsymbol{\alpha}$ be ℓ-bit rationals describing a transient $(M, \boldsymbol{\alpha})$-backoff process and let π be its limit probability vector. For every $\varepsilon > 0$, there exists $\beta > 0$, with $\log \frac{1}{\beta} = \operatorname{poly}(n, \ell, \log \frac{1}{\varepsilon})$, such that given any vector $\mathbf{r}'$ of ℓ'-bit rationals satisfying $\|\mathbf{r}' - \mathbf{r}\|_\infty \leq \beta$, a vector π' satisfying $\|\pi' - \pi\|_\infty \leq \varepsilon$ can be found in time $\operatorname{poly}(n, \ell, \ell', \log \frac{1}{\varepsilon})$.*

PROOF. Let $\mathbf{r}'$ be such that $\|\mathbf{r}' - \mathbf{r}\|_\infty \leq \beta$ (where β will be specified later). We will assume for notational simplicity that $r'_i \geq r_i$ for every i. (If this is not the case, then the vector $\mathbf{r}''$ given by $r''_i = r'_i + \beta$ does satisfy this property and still satisfies $\|\mathbf{r}'' - \mathbf{r}\|_\infty \leq 2\beta$. Thus the proof below with $\mathbf{r}'$ replaced by $\mathbf{r}''$ and β by 2β will work for the general case.)

Let D, D_α and X be as in the proof of Lemma 4.37. Define D', D'_α and X' analogously. Thus, 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

$$D' = \begin{pmatrix} D'_\alpha & X' \\ 0 & 0 \end{pmatrix},$$

where D'_α is irreducible. Notice first that $X' = X$, since if $\alpha_j = 1$, then for each i we have $D_{ij} = D'_{ij} = M_{ij}(1-\alpha_i)$. Recall that our goal is to approximate the maximal left eigenvector π of D, such that $\|\pi\|_1 = 1$. Write $\pi = 1/(1+\ell_B)[\pi_A \pi_B]$, where π_A is a left eigenvector of D_α with $\|\pi_A\|_1 = 1$, $\pi_B = \pi_A X$ and $\ell_B = \|\pi_B\|_1$. We will show how to compute π'_A, π'_B such that $\|\pi'_A\|_1 = 1$, $\|\pi'_A - \pi_A\|_\infty \le \varepsilon/(n+1)$ and $\|\pi'_B - \pi_B\|_\infty \le \varepsilon/(n+1)$. It follows then that if we set $\pi' = 1/(1+\|\pi'_B\|_1)[\pi'_A \pi'_B]$, then

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

as desired. [The term $|\ell_B - \|\pi'_B\|_1|$ is an upper bound on $\|\pi'_A/(1+\ell_b) - \pi'_A/(1+\|\pi'_B\|_1)\|_\infty$, and also on $\|\pi'_B/(1+\ell_b) - \pi'_B/(1+\|\pi'_B\|_1)\|_\infty$.]

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

Thus, below we show how to find π'_A that closely approximates π_A, specifically satisfying $\|\pi'_A - \pi_A\|_\infty \le \varepsilon/(n(n+1))$. To do so, we will use the matrix D'.

We now show that the entries of D' are close to those of D, using the fact that $0 \le r'_k - r_k \le \beta$. Note that

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

Thus to upper bound this difference, we need a 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 $\alpha_j \ge 2^{-\ell}$. 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 $r_k \le 1 - 2^{-2n\ell}$. This is true, since the (M, α)-backoff process is irreducible and hence there is a path consisting only of forward steps that goes from k to j, and this path has probability at least $2^{-2n\ell}$; once we push j onto the history stack, it will never be revoked. Further, by the irreducibility of the $(M, \boldsymbol{\alpha})$-backoff process, there must exist k_0 such that $M_{jk_0} > 0$ and $r_{k_0} \le 1 - 2^{-2n\ell}$. Now $M_{jk_0} \ge 2^{-\ell}$. Since $\sum_k M_{jk} = 1$, we have $\sum_{k \ne k_0} M_{jk} r_k + M_{jk_0} \le 1$, that is, $\sum_k M_{jk} r_k - M_{jk_0} r_{k_0} + M_{jk_0} \le 1$. So $\sum_k M_{jk} r_k \le 1 - M_{jk_0}(1 - r_{k_0}) \le 1 - 2^{-(2n+1)\ell}$. Hence, $1 - (1-\alpha_j)\sum_k M_{jk} r_k$ is lower bounded by $2^{-(2n+1)\ell}$. Thus we conclude that

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

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 $2^{-(2\ell+1)n}$. To see this, first note that every nonzero entry of D_α is at least $2^{-2\ell}$. Next consider a sequence $i_0 = i, i_1, i_2, \ldots, i_\ell = j$ 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 $2^{-n(2\ell+1)}$. Thus B is a positive matrix and we are interested in computing a close approximation to its left eigenvector π_A.

Next we show that $B' = \left(\frac{1}{2}(I + D'_\alpha)\right)^n$ is a close enough approximation to B. Note that since $\max_{ij}|D_{ij} - D'_{ij}| \le 2^{(4n+2)\ell}\beta$, we have $\max_{ij}|B'_{ij} - B_{ij}| \le (1 + 2^{(4n+2)\ell}\beta)^n - 1)$, which may be bounded from above by $(2^n \cdot 2^{(4n+2)\ell})\beta$ provided $\beta \le 2^{-(4n+2)\ell}$. [This follows from the fact that if $x \le 1$, then $(1+x)^n - 1 \le 2^n x$, which we can see by considering the binomial expansion of $(1+x)^n$ and noting that the sum of the coefficients is 2^n.]

Now let π'_A be any vector satisfying $\|\pi'_A - \pi'_A B'\|_\infty \le 2^{n+\ell(4n+2)}\beta$ 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 Theorem B.1 below to $B^T, (B')^T, \pi_A$ and π'_A with $\gamma = 2^{-n(2\ell+1)}$, $\varepsilon = \delta = 2^{n+\ell(4n+2)}\beta$ yields $\|\pi'_A - \pi_A\|_\infty \le \beta 2^{O(n\ell)}$. Thus setting $\beta = \varepsilon^2/2^{-\Omega(n\ell)}$ suffices to get π'_A to be an $\varepsilon/(n(n+1))$ close approximation to π_A. This concludes the proof. □

The rest of this section is devoted to the proof of the following theorem. As pointed out earlier, this theorem is due to Leonid Gurvits, and its proof is included here for completeness.

THEOREM B.1 [5]. *Let B, C be $n \times n$ matrices and $\mathbf{x}, \mathbf{y}$ be n-dimensional vectors satisfying the following conditions*:

(i) $B_{ij} \ge \gamma > 0$ *for every* i, j. *Further*, $\rho(B) = 1$. [*Recall that* $\rho(B)$ *is the maximal eigenvalue of* B.]

(ii) $|C_{ij} - B_{ij}| < \delta$ *for every* i, j.

(iii) $\|\mathbf{x}\|_1 = 1$ *and* $B\mathbf{x} = \mathbf{x}$.

(iv) $\|\mathbf{y}\|_1 = 1$ *and* $\|C\mathbf{y} - \mathbf{y}\|_\infty \le \varepsilon$.

Then $\|\mathbf{x} - \mathbf{y}\|_\infty \le (\varepsilon + \delta)/\gamma^3$, *provided* $\varepsilon + \delta \le \frac{\gamma}{2}$.

To prove the above theorem, we need to introduce some new definitions. In particular, a "projective norm" on vectors introduced by Hilbert, a norm on positive matrices induced by Hilbert's projective norm and a theorem of Birkhoff bounding the matrix norm play a crucial role in the proof of Theorem B.1. We introduce this background material next.

DEFINITION B.2 For n-dimensional positive vectors $\mathbf{x}$ and $\mathbf{y}$ the Hilbert projective distance between $\mathbf{x}$ and $\mathbf{y}$, denoted $d(\mathbf{x}, \mathbf{y})$, is defined to be

$$\ln \frac{\beta}{\alpha}, \text{ where } \alpha = \min_i \left\{ \frac{x_i}{y_i} \right\} \text{ and } \beta = \max_i \left\{ \frac{x_i}{y_i} \right\}.$$

It may be verified that for every $\gamma_1, \gamma_2 > 0$, it holds that $d(\mathbf{x}, \mathbf{y}) = d(\gamma_1 \cdot \mathbf{x}, \gamma_2 \cdot \mathbf{y})$, and thus $d(\cdot, \cdot)$ is invariant under scaling of vectors. Further, the projective norm satisfies the three properties of metrics (on the projective space), namely (1) nonnegativity; that is, $d(\mathbf{x}, \mathbf{y}) \ge 0$ with equality holding if and only if $\mathbf{x} = \mathbf{y}$, (2) symmetry; that is, $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ and (3) the triangle inequality; that is, $d(\mathbf{x}, \mathbf{y}) \le d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$. In the following lemma, we relate the ℓ_∞-distance between two positive unit vectors in the ℓ_1-norm with the projective distance between the two.

LEMMA B.3 *Let* $\mathbf{x}, \mathbf{y}$ *be positive vectors. Then the following hold*:

(i) $d(\mathbf{x}, \mathbf{y}) \le 3\|\mathbf{x} - \mathbf{y}\|_\infty / \min_i\{y_i\}$, *provided* $\|\mathbf{x} - \mathbf{y}\|_\infty \le (\min_i\{y_i\})/2$.
(ii) *If* $\|\mathbf{x}\|_1 = 1$ *and* $\|\mathbf{y}\|_1 = 1$, *then* $\|\mathbf{x} - \mathbf{y}\|_\infty \le d(\mathbf{x}, \mathbf{y})$.

PROOF. Let $\varepsilon = \|\mathbf{x} - \mathbf{y}\|_\infty$ and $\gamma = \min_i\{y_i\}$. For Part (i), note that $x_i/y_i \le 1 + |x_i - y_i|/y_i \le 1 + \frac{\varepsilon}{\gamma}$. (The first inequality holds by considering two cases: if $x_i \le y_i$, then the left-hand side is at most 1; if $x_i > y_i$, then the right-hand side equals x_i/y_i. Similarly, considering the two cases $x_i \le y_i$ and $x_i > y_i$, we obtain $x_i/y_i \ge 1 - |x_i - y_i|/y_i \ge 1 - \frac{\varepsilon}{\gamma}$. Thus $d(\mathbf{x}, \mathbf{y}) \le \ln((1 + \frac{\varepsilon}{\gamma})/(1 - \frac{\varepsilon}{\gamma})) = \ln(1 + \frac{\varepsilon}{\gamma}) + \ln(1/(1 - \frac{\varepsilon}{\gamma}))$. Using the inequality $\ln(1+z) \le z$, we see that the first term is at most $\frac{\varepsilon}{\gamma}$. For the second term, we use the fact that $(1/(1-z)) \le 1+2z$, if $z \le \frac{1}{2}$. Combined with the monotonicity of the natural logarithm, we get that $\ln(1/(1 - \frac{\varepsilon}{\gamma})) \le \ln(1 + 2\frac{\varepsilon}{\gamma}) \le 2\frac{\varepsilon}{\gamma}$, where the first inequality holds provided $\varepsilon \le \gamma/2$. If follows that $d(\mathbf{x}, \mathbf{y}) \le 3\frac{\varepsilon}{\gamma}$, provided $\varepsilon \le \gamma/2$.

For Part (ii), let i_0 be such that $|x_{i_0} - y_{i_0}| = \varepsilon$. Assume without loss of generality that $x_{i_0} = y_{i_0} + \varepsilon$. Since $\sum_j x_j = 1$, we have $x_{i_0} \le 1$. Therefore $x_{i_0}/(x_{i_0} - \varepsilon) \ge 1/(1 - \varepsilon)$ (as we see by clearing the denominators in the inequality), that is $x_{i_0}/y_{i_0} \ge 1/(1 - \varepsilon)$. Thus $\max_i\{x_i/y_i\} \ge 1/(1 - \varepsilon)$. Since $\sum_j x_j = \sum_j y_j$, there must exist an index i_1 such that $y_{i_1} \ge x_{i_1}$. Thus $\min_i\{x_i/y_i\} \le 1$. Putting the above together, we get $d(\mathbf{x}, \mathbf{y}) \ge \ln(1/(1 - \varepsilon)) \ge \varepsilon$. □

The Hilbert projective distance between vectors induces a natural norm on positive matrices, as defined below.

DEFINITION B.4 For a positive square matrix A, define the projective norm of A, denoted $\rho_{\mathrm{H}}(A)$, to be

$$\rho_{\mathrm{H}}(A) = \sup_{\mathbf{x},\mathbf{y}>0}\left\{\frac{d(A\mathbf{x}, A\mathbf{y})}{d(\mathbf{x}, \mathbf{y})}\right\}.$$

It turns out that the projective norm of every positive matrix is strictly smaller than 1. This can be shown using a theorem of Birkhoff that we will state shortly. First we need one more definition related to positive matrices.

DEFINITION B.5 For a positive square matrix A, define the diameter of A, denoted $\Delta(A)$, to be

$$\Delta(A) = \sup_{\mathbf{x},\mathbf{y}>0}\{d(A\mathbf{x}, A\mathbf{y})\}.$$

Birkhoff's theorem below relates the projective norm of a matrix to its diameter. In particular it shows that if the diameter of a matrix is bounded, then its projective norm is strictly smaller than 1.

THEOREM B.6 [2]. *For every positive square matrix A,*

$$\rho_{\mathrm{H}}(A) = \tanh(\Delta(A)/4).$$

Recall that $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$, and so $\tanh(x) < 1$ for every x. In the following lemma it is shown that the diameter of every positive matrix is bounded, and thus every positive matrix has a projective norm less than 1.

LEMMA B.7 *For a positive square matrix A satisfying $\rho(A) = 1$ and $A_{ij} \geq \gamma > 0$, it is the case that $\rho_{\mathrm{H}}(A) \leq 1 - \gamma^2$.*

PROOF. Let $\mathbf{z}$ be the maximal right eigenvector of A normalized to satisfy $\|\mathbf{z}\|_1 = 1$. (Note $\mathbf{z}$ is positive by the Perron–Frobenius theorem.) Let $\widetilde{A} = D^{-1}AD$, where D is the diagonal matrix with ith diagonal entry being z_i. We bound $\rho_{\mathrm{H}}(A)$ in four steps showing: (1) $\rho_{\mathrm{H}}(A) = \rho_{\mathrm{H}}(\widetilde{A})$, (2) $\widetilde{A}$ is row-stochastic (i.e., its rows sum to one), (3) $\widetilde{A}_{ij} \geq \gamma^2$ and (4) $\Delta(\widetilde{A}) \leq 1 - \min_{i,j}\{\widetilde{A}_{ij}\}$ for each row-stochastic matrix $\widetilde{A}$.

For step (1), note first that by the definition of the projective distance, we have $d(D_1\mathbf{x}D_2, D_1\mathbf{y}D_2) = d(\mathbf{x}, \mathbf{y})$ for each pair $\mathbf{x}, \mathbf{y}$ of vectors and each pair D_1, D_2 of positive diagonal matrices. As a consequence, we find that for each choice of a positive matrix A and positive diagonal matrices D_1 and D_2, we have $\rho_{\mathrm{H}}(A) = \rho_{\mathrm{H}}(D_1AD_2)$. Setting $D_1 = D^{-1}$ and $D_2 = D$ yields $\rho_{\mathrm{H}}(A) = \rho_{\mathrm{H}}(\widetilde{A})$.

For step (2), we need to verify that $\sum_j \widetilde{A}_{ij} = 1$ for every i. Note that $\widetilde{A}_{ij} = A_{ij} \cdot z_j/z_i$. Summing, we get $\sum_j A_{ij}(z_j/z_i) = (1/z_i)\sum_j A_{ij}z_j = 1$, where the last equality uses the fact that $A\mathbf{z} = \mathbf{z}$.

For step (3), we need to verify that $A_{ij}(z_j/z_i) \geq \gamma^2$. Since we know $A_{ij} \geq \gamma$, it suffices to show that $z_j \geq \gamma$ and $z_i \leq 1$. For the former, note that $z_j = \sum_k A_{jk} z_k \geq \sum_k \gamma z_k = \gamma$ (since $A_{jk} \geq \gamma$ and $\|\mathbf{z}\|_1 = 1$). For the latter, we use $z_i \leq \sum_k z_k = 1$. Thus we get $\tilde{A}_{ij} \geq \gamma^2$.

Finally for step (4), let $\mu = \min_{i,j}\{\tilde{A}_{ij}\}$. Assume that $\mathbf{x}$ and $\mathbf{y}$ are vectors of ℓ_1-norm 1. Then $\mu \leq (\tilde{A}\mathbf{x})_i \leq 1$ and $\mu \leq (\tilde{A}\mathbf{y})_i \leq 1$. Hence, $\mu \leq (\tilde{A}\mathbf{x})_i/(\tilde{A}\mathbf{y})_i \leq \frac{1}{\mu}$. Thus for every $\mathbf{x}$ and $\mathbf{y}$ of ℓ_1-norm 1, we have $d(\tilde{A}\mathbf{x}, \tilde{A}\mathbf{y}) \leq -2\ln\mu$. Hence

$$\Delta(\tilde{A}) = \sup_{\mathbf{x},\mathbf{y}>0}\{d(\tilde{A}\mathbf{x}, \tilde{A}\mathbf{y})\} = \sup_{\mathbf{x},\mathbf{y}>0,\, \|\mathbf{x}\|_1=\|\mathbf{y}\|_1=1}\{d(\tilde{A}\mathbf{x}, \tilde{A}\mathbf{y})\} \leq -2\ln\mu,$$

where the second equality holds since the projective distance is invariant with respect to scaling of the arguments. From Theorem B.6 and the fact that $\tanh(x) \leq 1 - e^{-2x}$, we get $\rho_{\rm H}(\tilde{A}) = \tanh(\Delta(\tilde{A})/4) \leq 1 - e^{-\Delta(\tilde{A})/2} \leq 1 - e^{-\ln\mu} = 1 - \mu$. □

Next we derive an easy corollary of Lemma B.7.

LEMMA B.8 *If A is a positive matrix with maximal right eigenvector $\mathbf{x}$, then* $\lim_{k\to\infty}\{d(A^k\mathbf{y}, \mathbf{x})\} = 0$ *for every positive vector $\mathbf{y}$.*

PROOF. Assume without loss of generality that $\rho(A) = 1$, and $A\mathbf{x} = \mathbf{x}$ (since A may be scaled without affecting its projective properties). Note that $d(A^k\mathbf{y}, A^k\mathbf{x}) \leq \rho_{\rm H}(A)d(A^{k-1}\mathbf{y}, A^{k-1}\mathbf{x})$ by the definition of the projective norm $\rho_{\rm H}(\cdot)$. Since $A^k\mathbf{x} = \mathbf{x}$, we get that $d(A^k\mathbf{y}, \mathbf{x}) \leq \rho_{\rm H}(A)^k d(\mathbf{y}, \mathbf{x})$. Since $\rho_{\rm H}(A) < 1$, we have $d(A^k\mathbf{y}, \mathbf{x})$ tends to 0 as $k \to \infty$. □

The next lemma shows that if $A\mathbf{y}$ is close to $\mathbf{y}$ for a positive matrix A with maximal eigenvalue 1 and positive vector $\mathbf{y}$, then $\mathbf{y}$ is close to the maximal eigenvector of A, where closeness is under the projective norm.

LEMMA B.9 *For a positive square matrix A, with maximal right eigenvector $\mathbf{x}$, if $\mathbf{y}$ satisfies $d(A\mathbf{y}, \mathbf{y}) \leq \varepsilon$, then $d(\mathbf{y}, \mathbf{x}) \leq \varepsilon/(1 - \rho_{\rm H}(A))$.*

PROOF. Again, we assume that $\rho(A) = 1$, to simplify the notation. Then from Lemma B.8 we have $\lim_{k\to\infty}\{d(A^k\mathbf{y}, \mathbf{x})\} = 0$. Thus, using triangle inequality on the projective distance we get $d(\mathbf{y}, \mathbf{x}) \leq \sum_{k=0}^{\infty} d(A^k\mathbf{y}, A^{k+1}\mathbf{y})$. But $d(A^k\mathbf{y}, A^{k+1}\mathbf{y}) \leq \rho_{\rm H}(A)d(A^{k-1}\mathbf{y}, A^k\mathbf{y}) \leq \rho_{\rm H}(A)^k d(\mathbf{y}, A\mathbf{y}) \leq \rho_{\rm H}(A)^k\varepsilon$. Thus, we have $d(\mathbf{y}, \mathbf{x}) \leq \sum_{k=0}^{\infty}\rho_{\rm H}(A)^k\varepsilon = \varepsilon/(1 - \rho_{\rm H}(A))$. □

PROOF OF THEOREM B.1. By Conditions (2) and (4) of the hypothesis we get

$$\|B\mathbf{y} - \mathbf{y}\|_\infty \leq \|(B - C)\mathbf{y}\|_\infty + \|C\mathbf{y} - \mathbf{y}\|_\infty \leq \delta + \varepsilon. \tag{14}$$

Applying Part (1) of Lemma B.3, where the roles of $\mathbf{x}$, $\mathbf{y}$ of the lemma are played here by $\mathbf{y}$, $B\mathbf{y}$, respectively, we get $d(\mathbf{y}, B\mathbf{y}) \leq 3(\varepsilon+\delta)/\gamma$. (Note the necessary condition for the application of Lemma B.3 follows from the condition

$\varepsilon+\delta \leq \gamma/2$. Applying Lemma B.9, where the roles of A, $\mathbf{x}$, $\mathbf{y}$ of the lemma are played here by B, $\mathbf{x}$, $\mathbf{y}$, respectively, we get $d(\mathbf{y}, \mathbf{x}) \leq 3(\varepsilon+\delta)/(\gamma\cdot(1-\rho_{\mathrm{H}}(B)))$. By Lemma B.7 we have $\rho_{\mathrm{H}}(B) \leq 1-\gamma^2$, and thus $d(\mathbf{y}, \mathbf{x}) \leq (\varepsilon+\delta)/\gamma^3$. Applying part (ii) of Lemma B.3 to vectors $\mathbf{x}$ and $\mathbf{y}$, we get $\|\mathbf{x}-\mathbf{y}\|_\infty \leq d(\mathbf{x}, \mathbf{y}) \leq (\varepsilon+\delta)/\gamma^3$. □

Acknowledgment. An earlier version of this paper contained a flawed proof of Theorem B.1. We thank Leonid Gurvits for pointing out the error, for suggesting an alternate proof and for permission to include his proof in this paper.

REFERENCES

[1] BARRETT, R., MAGLIO, P. P. and KELLEM, D. C. (1997). How to personalize the web. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '97)* 75–82. Atlanta.
[2] BIRKHOFF, G. (1957). Extensions of Jentzsch's theorem. *Trans. Amer. Math. Soc.* **85** 219–227.
[3] CHARIKAR, M., KUMAR, S. R., RAGHAVAN, P., RAJAGOPALAN, S. and TOMKINS, A. (1999). On targeting Markov segments. *Proceedings of the ACM Symposium on Theory of Computing, 1999* 99–108. Atlanta.
[4] GROTSCHEL, M., LOVASZ, L. and SCHRIJVER, A. (1987). *Geometric Algorithms and Combinatorial Optimizations.* Springer, Berlin.
[5] GURVITS, L. (2000). Personal communication.
[6] HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1964). *Inequalities*. Cambridge Univ. Press.
[7] HORN, R. A. and JOHNSON, C. A. (1985). *Matrix Analysis.* Cambridge Univ. Press.
[8] KARLIN, S. and TAYLOR, H. (1975). *A First Course in Stochastic Processes.* Academic Press, New York.
[9] KEMENY, J. G., SNELL, J. L. and KNAPP, A. W. (1966). *Denumerable Markov Chains*. Van Nostrand, Princeton, New Jersey.
[10] LIEBERMAN, H. (1995). An agent that assists web browsing. *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI '95)* 924–929.
[11] MINC, H. (1988). *Nonnegative Matrices*. Wiley, New York.
[12] MOTWANI, R. and RAGHAVAN, P. (1994). *Randomized Algorithms*. Cambridge Univ. Press.
[13] SELKER, T. (1994). COACH: a teaching agent that learns. *Communications of the ACM* **37** 92–99.

R. FAGIN
S. RAJAGOPALAN
A. TOMKINS
IBM ALMADEN RESEARCH CENTER
DEPARTMENT K53/B2
650 HARRY ROAD
SAN JOSE, CALIFORNIA 95120
E-MAIL: {fagin,sridhar,tomkins}@almaden.ibm.com

A. R. KARLIN
DEPARTMENT OF COMPUTER SCIENCE
BOX 352350
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
E-MAIL: karlin@cs.washington.edu

J. KLEINBERG
DEPARTMENT OF COMPUTER SCIENCE
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853
E-MAIL: kleinber@cs.cornell.edu

P. RAGHAVAN
VERITY, INC.
892 ROSS DRIVE
SUNNYVALE, CALIFORNIA 94089
E-MAIL: pragh@verity.com

R. RUBINFELD
NECI
4 INDEPENDENCE WAY
PRINCETON, NEW JERSEY 08540
E-MAIL: ronitt@research.nj.nec.com

M. SUDAN
MIT LABORATORY FOR COMPUTER SCIENCE
545 TECHNOLOGY SQUARE NE43-307
CAMBRIDGE, MASSACHUSETTS 02139
E-MAIL: madhu@mit.edu