The move-to-front rule for self-organizing lists with Markov dependent requests

[short title: Self-organizing search with Markov requests]

by Robert P. Dobrow and James Allen Fill[1]

The Johns Hopkins University

Abstract

We consider the move-to-front self-organizing linear search heuristic where the sequence of record requests is a Markov chain. Formulas are derived for the transition probabilities and stationary distribution of the permutation chain. The spectral structure of the chain is presented explicitly. Bounds on the discrepancy from stationarity for the permutation chain are computed in terms of the corresponding discrepancy for the request chain, both for separation and for total variation distance.

1 Introduction and summary

A collection of n records is arranged in a sequential list. Associated with the ith record is a weight r_i measuring the long-run frequency of its use. We assume that each $r_i > 0$ and normalize so that $\sum r_i = 1$. At each unit of time, item i is removed from the list with probability r_i and replaced at the front of the list. This gives a Markov chain on the permutation group S_n.

If we assume that items are requested independently of all other requests, this model for dynamically organizing a sequential file is known as the

[1]Research for both authors supported by NSF grant DMS-9311367.

[2]*AMS 1991 subject classifications.* Primary 60J10; secondary 68P10, 68P05.

[3]*Keywords and phrases.* Markov chains, self-organizing search, move-to-front rule, convergence to stationarity, separation, total variation distance, coupling.

move-to-front (MTF) heuristic and has been studied extensively for over 20 years. Background references include Rivest (1976), Bitner (1979), Hendricks (1989), Fill (1993), and Diaconis (1993). In the case when all the weights are equal the model corresponds to a card-shuffling scheme known as the random-1-to-top shuffle; see Diaconis et al. (1992) for a thorough analysis in this case.

One objection to this model is that in practice record requests tend to exhibit "locality of reference." That is, frequencies of access over the short run may differ quite substantially from those over the long run. Knuth (1973) and Bentley and McGeoch (1985), among others, have noted that MTF tends to work even better in practice than predicted from the i.i.d. model. Knuth cites computational experiments involving compiler symbol tables and notes that typically "successive searches are not independent (small groups of keys tend to occur in bunches)."

Konnecker and Varol (1981) proposed modeling the request sequence along Markovian or autocorrelative lines. Lam et al. (1984) formally set up the Markovian model for self-organizing search and obtained a formula for the asymptotic average cost of searching for a record (i.e., for stationary expected search cost). Phatarfod and Dyte (1993) have derived the eigenvalues of the transition matrix for this *Markov move-to-front* (MMTF) model.

In this paper we assume the Markovian model. We call the Markov chain corresponding to the sequence of record requests the *request chain.* We derive explicit formulas for the transition probabilities and stationary distribution of Markov move-to-front. We also study convergence to stationarity for MMTF and obtain bounds on separation and total variation distance of the MMTF chain from its stationary distribution in terms of the discrepancy from stationarity for the request chain.

After setting notation and addressing preliminary issues we discuss models for the request chain in Section 2. In Section 3, the k-step transition probabilities for MMTF are derived. In Section 4, we give three formulas for the stationary distribution. One has a direct probabilistic interpretation; the other two are more suited for numerical calculation in terms of the request matrix or its time reversal. In Section 5 we give the spectral structure of the MMTF chain. In the final two sections we analyze the speed of convergence of MMTF to its stationary distribution. In Section 6 we treat separation and obtain bounds in terms of the separation of the request chain. In Section 7

we treat total variation distance. In brief, our approach there is as follows. Variation distance is bounded above by the tails of any coupling time. We couple two copies of the MMTF chain by first coupling the corresponding request chains and then use the standard coupling for MTF, namely, wait until all but one of the records has been requested at least once.

2 Preliminaries and discussion of models

Let $\sigma \in S_n$ represent an ordered list of records, with σ_k denoting the record at the kth position in the list. Let $r_1, \ldots, r_n$ be a sequence of probabilities (weights), with the interpretation that record i is requested with long-run frequency r_i. We assume that all the probabilities are strictly positive. Let $[n] := \{1, \ldots, n\}$.

Let R be the $n \times n$ transition matrix for the request chain. Thus $R(i, j)$ is the probability of accessing record j given that the previous request was for record i. In the case of independent requests the rows of R are identical and equal to $(r_1, \ldots, r_n)$. We will denote such a request matrix by R_0 and refer to MMTF with such a request matrix as the *i.i.d. case* or as MTF.

Several authors (e.g., Lam et al. (1984) and Kapoor and Reingold (1991)) have considered the model

$$R := (1 - \alpha)R_0 + \alpha I_n$$

for the request chain R, where I_n is the $n \times n$ identity matrix and $\alpha \in [0, 1]$. The results of this paper can be easily applied to this case using the fact that

$$Q^k = \sum_{j=0}^{k} \binom{k}{j} (1 - \alpha)^j \alpha^{k-j} Q_0^j, \quad k \geq 0,$$

where Q_0 is the transition matrix for MTF and Q is the transition matrix for MMTF.

A generalization of this model which seems to capture locality of reference reasonably well is a mixture of the i.i.d. chain and a birth-and-death chain, that is,

$$R := (1 - \alpha)R_0 + \alpha B, \tag{1}$$

where B is a birth-and-death transition matrix. Unfortunately, we do not know much in the way of neat formulas for this case. The analysis leads to

quite difficult problems involving taboo probabilities and covering times for Markov chains. We shall, however, treat the extreme case when $\alpha = 1$ in (1), that is, the *birth-and-death case* $R = B$. This example may not be so realistic in that it exhibits *too much* locality of reference. However, it does provide an interesting and non-trivial request chain for which fairly complete results can be obtained.

There are several ways to model MMTF. Lam et al. (1984) model the chain on the state space $S_n \times [n]$, where (σ, i) denotes (present configuration, next request). This approach has the advantage of being able to handle other self-organizing schemes besides move-to-front. It is clear, as Phatarfod and Dyte (1993) point out, that the state space can also be specified as (present configuration, last request). Since for MMTF the configuration itself incorporates information about the last request—the last request can be read from σ as σ_1—we can and do take S_n as the state space for MMTF.

Let Q be the transition matrix for MMTF. Then

$$Q(\pi, \sigma) = \begin{cases} R(\pi_1, \sigma_1), & \text{if } \sigma = \pi \circ (k \cdots 1), \text{ where } k = \pi^{-1}(\sigma_1), \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

In all that follows we require that the R-chain be ergodic, which implies that its stationary distribution is unique and strictly positive. Note that this does *not* imply that the permutation chain is ergodic; see Section 4 for further discussion.

3 Transition probabilities

Before proceeding to the main result of this section (Theorem 1) we establish some terminology and notation. Say that a permutation σ has a *descent* at position i if $\sigma_i > \sigma_{i+1}$. Denote by $L(\sigma)$ the last descent position for σ, with $L(\mathrm{id}) := 0$, where id denotes the identity permutation. Note that $L(\pi^{-1}\sigma)$ is the minimum number of entries of π that must be moved to the front to obtain σ.

In what follows, the time-reversal of the R chain will arise naturally. Let $\widetilde{R}$ be the transition matrix of the time-reversed chain; that is, $\widetilde{R}(x, y) := r_y R(y, x)/r_x$, where $\mathbf{r} = (r_1, \ldots, r_n)$ is the stationary distribution of the R chain.

For a vector or permutation σ of length at least m, we write $\sigma_{\to m}$ to denote the m-element vector consisting of the first m elements of σ. Notation such as $y_{t_{\to m}}$ is shortand for $(y_{t_1}, \ldots, y_{t_m})$. We write $\sigma[m]$ to denote the unordered set $\{\sigma_1, \ldots, \sigma_m\}$. For a matrix A and set S, let A_S denote the principal submatrix of A determined by the rows and columns identified by S.

Let $\widetilde{X} = (\widetilde{X}_k)_{k \geq 0}$ be a Markov chain with transition matrix $\widetilde{R}$. We use the notation $P_x(\cdot)$ for conditional probability given that $\widetilde{X}_0 = x$ and $P_{\mathbf{r}}(\cdot)$ for probability with respect to the $\widetilde{X}$-chain started in its stationary distribution $\mathbf{r}$.

We next define the *partial cover time* $\widetilde{C}_m$, the first time at which $\widetilde{X}$ has visited m distinct states. Formally, we can define $\widetilde{C}_m$ as follows:

$$\widetilde{C}_m := \inf\{k \geq 0 : |\widetilde{X}[k]| = m\},$$

recalling that $\widetilde{X}[k]$ denotes the unordered set $\{\widetilde{X}_1, \ldots, \widetilde{X}_k\}$. In particular, $\widetilde{C}_1 = 0$, $\widetilde{C}_n = \widetilde{C}$ is the usual cover time for $\widetilde{X}$, and $\widetilde{C}_m = \infty$ for $m > n$.

A key observation for analyzing the MTF chain, utilized by Fill (1993) in obtaining an exact formula for the k-step transition probabilities, is that one can read off the sequence of last requests of distinct records from the order of the list. That is, if $\sigma \in S_n$ is the final order of the list, then σ_1 must have been requested last, σ_2 is the penultimate distinct record to have been requested, etc. We will make use of this observation in the proof of the following theorem.

Theorem 1 *Let $\pi, \sigma \in S_n$. Under MMTF,*

$$Q^k(\pi, \sigma) = \frac{1}{r_{\pi_1}} \sum_{m=L(\pi^{-1}\sigma)}^{n} P_{\mathbf{r}}[\widetilde{X}_{\widetilde{C}_{\to m}} = \sigma_{\to m},\, \widetilde{C}_m < k \leq \widetilde{C}_{m+1},\, \widetilde{X}_k = \pi_1] \quad (3)$$

for $k \geq 0$, with the convention that if $\pi = \sigma$, the summand for $m = 0$ in (3) equals r_{π_1} if $k = 0$ and vanishes otherwise.

Proof Let $Q^k(\pi, \sigma; m)$ denote the probability, starting in a fixed π, that k requests move exactly m distinct records to the front of the list and result in the permutation σ. Thus

$$Q^k(\pi, \sigma) = \sum_{m=0}^{n} Q^k(\pi, \sigma; m).$$

Note that $Q^k(\pi,\sigma;m)$ vanishes if $L(\pi^{-1}\sigma) > m$. If $m \geq 1$ and $L(\pi^{-1}\sigma) \leq m$, then by noting that the first m records must have their last requests occur in the order $\sigma_m, \ldots, \sigma_1$, and conditioning on the times of these requests, we find

$$Q^k(\pi,\sigma;m) = \sum_{j_{\to m}} \left[\prod_{v=1}^{m-1} \sum_{i=1}^{v} R(\sigma_{v+1},\sigma_i) R^{j_v}_{\sigma[v]}(\sigma_i,\sigma_v) \right] \sum_{i=1}^{m} R(\pi_1,\sigma_i) R^{j_m}_{\sigma[m]}(\sigma_i,\sigma_m), \tag{4}$$

where the outer sum is over all m-tuples $j_{\to m} = (j_1,\ldots,j_m)$ whose elements are nonnegative integers summing to $k-m$. The quantity $R^{j_v}_{\sigma[v]}(\sigma_i,\sigma_v)$ is the "taboo" probability of moving from state σ_i to σ_v in j_v steps while hitting only states in $\sigma[v]$.

Let $\rho = r_{\sigma_1}/r_{\pi_1}$. Passing to the time-reversed matrix, (4) equals

$$\begin{aligned}
& \sum_{j_{\to m}} \left[\prod_{v=1}^{m-1} \frac{r_{\sigma_v}}{r_{\sigma_{v+1}}} \sum_{i=1}^{v} \widetilde{R}^{j_v}_{\sigma[v]}(\sigma_v,\sigma_i) \widetilde{R}(\sigma_i,\sigma_{v+1}) \right] \frac{r_{\sigma_m}}{r_{\pi_1}} \sum_{i=1}^{m} \widetilde{R}^{j_m}_{\sigma[m]}(\sigma_m,\sigma_i) \widetilde{R}(\sigma_i,\pi_1) \\
= \; & \rho \sum_{j_{\to m}} \left[\prod_{v=1}^{m-1} \sum_{i=1}^{v} \widetilde{R}^{j_v}_{\sigma[v]}(\sigma_v,\sigma_i) \widetilde{R}(\sigma_i,\sigma_{v+1}) \right] \sum_{i=1}^{m} \widetilde{R}^{j_m}_{\sigma[m]}(\sigma_m,\sigma_i) \widetilde{R}(\sigma_i,\pi_1) \qquad (5) \\
= \; & \begin{cases} \rho P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to m}} = \sigma_{\to m},\, \widetilde{C}_m < k,\, \widetilde{C}_{m+1} = k,\, \widetilde{X}_k = \pi_1], & \text{if } \pi_1 \notin \sigma[m] \\ \rho P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to m}} = \sigma_{\to m},\, \widetilde{C}_m < k,\, \widetilde{C}_{m+1} > k,\, \widetilde{X}_k = \pi_1], & \text{if } \pi_1 \in \sigma[m] \end{cases} \\
= \; & \rho P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to m}} = \sigma_{\to m},\, \widetilde{C}_m < k \leq \widetilde{C}_{m+1},\, \widetilde{X}_k = \pi_1]. \qquad (6)
\end{aligned}$$

The result follows. ■

Remarks:

1. We can recapture Fill's (1993) formula for the k-step transition probabilities for MTF from (5). For fixed $\sigma \in S_n$, let $w_i := r_{\sigma_i}$ and $w_v^+ := \sum_{i=1}^v w_i$ with the convention that $w_0^+ := 0$. Note that in the i.i.d. case

$$\widetilde{R}^{j_v}_{\sigma[v]}(\sigma_v, \sigma_i) = (w_v^+)^{j_v - 1} w_i$$

for $1 \le i \le v$. Hence

$$\begin{aligned} & \left[\prod_{v=1}^{m-1} \sum_{i=1}^{v} \widetilde{R}^{j_v}_{\sigma[v]}(\sigma_v, \sigma_i) \widetilde{R}(\sigma_i, \sigma_{v+1})\right] \sum_{i=1}^{m} \widetilde{R}^{j_m}_{\sigma[m]}(\sigma_m, \sigma_i) \widetilde{R}(\sigma_i, \pi_1) \\ = & \left[\prod_{v=1}^{m-1} \sum_{i=1}^{v} (w_v^+)^{j_v - 1} w_i w_{v+1}\right] \sum_{i=1}^{m} (w_m^+)^{j_m - 1} w_i r_{\pi_1} \\ = & \; r_{\pi_1} \left(\prod_{v=2}^{m} w_v\right) \prod_{v=1}^{m} (w_v^+)^{j_v}. \end{aligned}$$

Thus from (5),

$$Q^k(\pi, \sigma) = \sum_{m = L(\pi^{-1}\sigma)}^{n} \left(\prod_{v=1}^{m} w_v\right) \sum_{j \to m} \prod_{v=1}^{m} (w_v^+)^{j_v}.$$

This calculation is the basis of the proof of Theorem 2.1 in Fill (1993).

2. For moderate n and a specific matrix R, the transition probabilities for MMTF can be computed directly using the formula in Theorem 1. The formula's usefulness can be appreciated, for instance, in the case of a 10-element list. The MMTF transition matrix is $3,628,800 \times 3,628,800$. The formula reduces the computations to calculations on matrices of size at most 10×10.

4 Stationary distribution

The main result of this section (Theorem 2) gives three representations of the stationary distribution for MMTF. One has a straightforward probabilistic interpretation, while the other two are more suited for numerical calculation in terms of the request matrix or its time reversal.

Theorem 2 *Let Q be the transition matrix for MMTF. Then, for any $\pi, \sigma \in S_n$,*

$$Q^k(\pi, \sigma) \to Q^\infty(\sigma) \text{ as } k \to \infty,$$

where

$$\begin{aligned} Q^\infty(\sigma) &= P_{\mathbf{r}}[\widetilde{X}_{\widetilde{C}_{\to n}} = \sigma] && (7) \\ &= r_{\sigma_n} \prod_{v=1}^{n-1} \left[\sum_{i=1}^{v} R(\sigma_{v+1}, \sigma_i)(I_{\sigma[v]} - R_{\sigma[v]})^{-1}(\sigma_i, \sigma_v) \right] && (8) \\ &= r_{\sigma_1} \prod_{v=1}^{n-1} \left[\sum_{i=1}^{v} (I_{\sigma[v]} - \widetilde{R}_{\sigma[v]})^{-1}(\sigma_v, \sigma_i) \widetilde{R}(\sigma_i, \sigma_{v+1}) \right], && (9) \end{aligned}$$

with I the $n \times n$ identity matrix.

Proof Let $k \to \infty$ in (3). All the terms on the right vanish in the limit except for the term with $m = n$, and a simple proof using the strong Markov property and the bounded convergence theorem shows that that term converges to the right side of (7).

Note that

$$P_{\mathbf{r}}[\widetilde{X}_{\widetilde{C}_{\to n}} = \sigma] = r_{\sigma_1} P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to (n-1)}} = \sigma_{\to (n-1)}].$$

Expanding this last expression in terms of the $\widetilde{R}$ matrix gives

$$\begin{aligned} Q^\infty(\sigma) &= r_{\sigma_1} \sum_{j_{\to (n-1)}} \prod_{v=1}^{n-1} \left[\sum_{i=1}^{v} \widetilde{R}^{j_v}_{\sigma[v]}(\sigma_v, \sigma_i) \widetilde{R}(\sigma_i, \sigma_{v+1}) \right] && (10) \\ &= r_{\sigma_1} \prod_{v=1}^{n-1} \left[\sum_{i=1}^{v} (I_{\sigma[v]} - \widetilde{R}_{\sigma[v]})^{-1}(\sigma_v, \sigma_i) \widetilde{R}(\sigma_i, \sigma_{v+1}) \right]. \end{aligned}$$

The outer sum in (10) is over all $(n-1)$-tuples $j_{\to (n-1)} = (j_1, \ldots, j_{n-1})$ whose elements are nonnegative integers. This proves (9), and (8) follows immediately from the relationship between R and $\widetilde{R}$.

■

Remark:

In the i.i.d. case, it follows from (7) that for $\sigma \in S_n$, $Q^\infty(\sigma)$ is the probability of obtaining the ordered list $(\sigma_1, \ldots, \sigma_n)$ by sampling without replacement from $\{\sigma_1, \ldots, \sigma_n\}$. Thus, in the notation $w_i = r_{\sigma_i}$ and $w_v^+ = \sum_{i=1}^v w_i$ introduced earlier,

$$Q^\infty(\sigma) = \frac{w_1 \cdots w_n}{\prod_{j=0}^{n-1}(1 - w_j^+)}.$$

As pointed out in Section 2, the ergodicity of the R-chain is insufficient for the Q-chain to be ergodic. However, there always exists a unique positive recurrent communication class RC, with $Q^\infty(\sigma) > 0$ if $\sigma \in RC$ and $Q^\infty(\sigma) = 0$ if σ is transient, that is, if $\sigma \notin RC$. The following lemma gives a necessary and sufficient condition for $\sigma \in RC$.

Lemma 4.1 *Let $\sigma \in S_n$. Then $Q^\infty(\sigma) > 0$ if and only if for each $v \in [n-1]$ there exists $k \geq 1$ with*

$$\widetilde{R}^k_{\sigma[v+1]}(\sigma_v, \sigma_{v+1}) > 0. \tag{11}$$

Proof Clearly, (11) holds for some $k \geq 1$ if and only if

$$\sum_{i=1}^{v} \left[\widetilde{R}^j_{\sigma[v]}(\sigma_v, \sigma_i)\widetilde{R}(\sigma_i, \sigma_{v+1})\right] > 0$$

for some $j \geq 0$. The lemma then follows from (10). ■

In the remainder of this section we consider the birth-and-death case introduced in Section 2. Let R be an ergodic chain with transitions

$$R(i,j) = \begin{cases} q_i, & \text{if } j = i-1 \\ 1 - q_i - p_i, & \text{if } j = i \\ p_i, & \text{if } j = i+1 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

for $1 \leq i, j \leq n$, with $q_1 = p_n = 0$. By using results for hitting times for birth-and-death chains one can compute $Q^\infty(\sigma)$ explicitly.

Theorem 3 *Let R be as defined in (12). Given $\sigma \in S_n$, a necessary and sufficient condition for $Q^\infty(\sigma) > 0$ is that $(\sigma_2, \ldots, \sigma_n)$ be an interleaving of*

the two sequences $(\sigma_1 - 1, \sigma_1 - 2, \ldots, 1)$ *and* $(\sigma_1 + 1, \sigma_1 + 2, \ldots, n)$. *For such a* σ,

$$Q^{\infty}(\sigma) = \frac{\hat{r}_{\sigma_1}}{\sum_{i=1}^{n} \hat{r}_i} \prod_{j=1}^{n-2} h(\sigma; j), \tag{13}$$

where $h(\sigma; j)$ *is defined to be*

$$\begin{cases} 1, & \text{if } \min \sigma[j] = 1 \text{ or } \max \sigma[j] = n \\ \frac{\gamma(\min \sigma[j]-1, \sigma_j - 1)}{\gamma(\min \sigma[j]-1, \sigma_{j+1}-1)}, & \text{if } \sigma_j < \sigma_{j+1},\ 1 < \min \sigma[j], \text{ and } \max \sigma[j] < n \\ \frac{\gamma(\sigma_j, \max \sigma[j])}{\gamma(\sigma_{j+1}, \max \sigma[j])}, & \text{if } \sigma_j > \sigma_{j+1},\ 1 < \min \sigma[j], \text{ and } \max \sigma[j] < n, \end{cases}$$

$$\gamma(a, b) := \sum_{i=a}^{b} \prod_{j=1}^{i} \frac{q_j}{p_j} \quad \text{for } 1 \le a \le b < n, \tag{14}$$

and

$$\hat{r}_j := \begin{cases} 1, & j = 1 \\ \frac{p_1 \cdots p_{j-1}}{q_2 \cdots q_j}, & 1 < j \le n. \end{cases} \tag{15}$$

Proof Let $\widetilde{T}_i$ be the first hitting time of state i for $\widetilde{X}$. That is,

$$\widetilde{T}_i := \inf\{k \ge 0 : \widetilde{X}_k = i\}. \tag{16}$$

Similarly define $\widetilde{T}_A$ for a subset A of the state space. For $1 \le j \le n-2$, let

$$\hat{h}(\sigma; j) := P_{\sigma_j}[\widetilde{T}_{\sigma_{j+1}} < \widetilde{T}_{\{\sigma_{j+2}, \ldots, \sigma_n\}}]. \tag{17}$$

Then

$$Q^{\infty}(\sigma) = r_{\sigma_1} \prod_{j=1}^{n-2} \hat{h}(\sigma; j).$$

A consequence of Lemma 4.1 is that $Q^{\infty}(\sigma) > 0$ if and only if $(\sigma_2, \ldots, \sigma_n)$ is an interleaving of the two sequences

$$(\sigma_1 - 1, \sigma_1 - 2, \ldots, 1) \text{ and } (\sigma_1 + 1, \sigma_1 + 2, \ldots, n).$$

For such σ,

$$\hat{h}(\sigma; j) = \begin{cases} P_{\sigma_j}[\widetilde{T}_{\sigma_{j+1}} < \widetilde{T}_{(\min \sigma[j])-1}], & \text{if } \sigma_j < \sigma_{j+1} \\ P_{\sigma_j}[\widetilde{T}_{\sigma_{j+1}} < \widetilde{T}_{(\max \sigma[j])+1}], & \text{if } \sigma_j > \sigma_{j+1} \end{cases}$$

with the natural convention that $\hat{h}(\sigma; j) = 1$ if $\min \sigma[j] = 1$ or $\max \sigma[j] = n$. It is elementary and well known that for birth-and-death chains

$$P_x[\widetilde{T}_b < \widetilde{T}_a] = \frac{\sum_{y=a}^{x-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}, \quad a < x < b,$$

where

$$\gamma_y := \frac{q_1 \cdots q_y}{p_1 \cdots p_y}.$$

Thus $\hat{h}(\sigma; j)$ equals $h(\sigma; j)$ in the statement of the theorem. The stationary distribution for birth-and-death chains is well known, giving

$$r_{\sigma_1} = \frac{\hat{r}_{\sigma_1}}{\sum_{i=1}^{n} \hat{r}_i},$$

with $\hat{r}$ defined at (15). ■

We now specialize to a simple case for which we can obtain completely explicit results. In the following corollary we treat the case of a request chain whose stationary distribution $\mathbf{r}$ is uniform on $[n]$. This is the same stationary distribution as the request chain for MTF with equal weights. But, as we'll see, the MMTF chain exhibits behavior quite different from that of the i.i.d. case.

Corollary 4.1 *Let the R-chain be the following simple symmetric random walk on $[n]$: For fixed $0 < p \le 1/2$, $p_i = q_{i+1} = p$ for $i = 1, \ldots, n-1$. Then for positive recurrent states $\sigma \in S_n$,*

$$Q^{\infty}(\sigma) = \frac{1}{n\alpha!} \prod_{1 \le j \le \alpha:\ \sigma_j \sim \sigma_{j+1}} j,$$

where $\alpha = \min\{\sigma^{-1}(1), \sigma^{-1}(n)\}$ and $a \sim b$ means a is adjacent to b, that is, $b \in \{a-1, a+1\}$.

Proof For simple symmetric random walk,

$$\gamma(a, b) = b - a + 1, \quad 1 \le a < b \le n. \tag{18}$$

Thus $h(\sigma; j)$ equals

$$\begin{cases} 1, & \text{if } \min \sigma[j] = 1 \text{ or } \max \sigma[j] = n \\ \frac{\sigma_j - \min \sigma[j] + 1}{\sigma_{j+1} - \min \sigma[j] + 1}, & \text{if } \sigma_j < \sigma_{j+1},\ 1 < \min \sigma[j], \text{ and } \max \sigma[j] < n \\ \frac{\max \sigma[j] - \sigma_j + 1}{\max \sigma[j] - \sigma_{j+1} + 1}, & \text{if } \sigma_j > \sigma_{j+1},\ 1 < \min \sigma[j], \text{ and } \max \sigma[j] < n. \end{cases} \tag{19}$$

Further, since σ is an interleaving,

$$\text{if } \left\{ \begin{array}{l} \sigma_j < \sigma_{j+1} \\ \sigma_j > \sigma_{j+1} \end{array} \right\} \text{ then } \left\{ \begin{array}{l} \sigma_{j+1} = \max \sigma[j] + 1 \\ \sigma_{j+1} = \min \sigma[j] - 1 \end{array} \right\}. \tag{20}$$

Also, if $\sigma_j < \sigma_{j+1}$ then

$$\left\{ \begin{array}{l} \sigma_j = \max \sigma[j] \\ \sigma_j = \min \sigma[j] \end{array} \right\} \text{ if } \left\{ \begin{array}{l} \sigma_j \sim \sigma_{j+1} \\ \sigma_j \not\sim \sigma_{j+1} \end{array} \right\} \tag{21}$$

and similarly if $\sigma_j > \sigma_{j+1}$ with the roles of $\max \sigma[j]$ and $\min \sigma[j]$ reversed. From (19), (20), and (21) it follows that

$$h(\sigma; j) = \begin{cases} j/(j+1), & \text{if } \sigma_j \sim \sigma_{j+1} \\ 1/(j+1), & \text{if } \sigma_j \not\sim \sigma_{j+1}. \end{cases} \tag{22}$$

The result is a consequence of (22) and the fact that $\hat{r}_j = 1$ for $1 \le j \le n$. ■

Remarks:

1. In the setting of Corollary 4.1, $Q^\infty(\sigma)$ is maximized by $\sigma = \text{id}$ and $\sigma = \text{rev}$ (and only these), with $Q^\infty(\max) = 1/n$; $Q^\infty(\sigma)$, when positive, is minimized by exactly two permutations σ (e.g., by

$$\sigma = \left(\frac{n}{2}, \frac{n}{2} + 1, \frac{n}{2} - 1, \frac{n}{2} + 2, \frac{n}{2} - 2, \ldots, 1, n \right),$$

if n is even), with $Q^\infty(\min) = 1/n!$. There are $2^{n-1} \ll n!$ permutations $\sigma \in S_n$ with $Q^\infty(\sigma) > 0$.

2. With the simple symmetric random walk request chain discussed above, MMTF behaves quite differently from MTF with equal weights, even though both request chains have the same stationary distribution. This is underscored by the magnitude of stationary expected search cost ESC. In the random walk case, using the main result of Lam et al. (1984) one finds ESC $= 1 + 2p(H_n - 1)$, where $H_n = \sum_{k=1}^{n} k^{-1}$. In the MTF case, ESC $= (n+1)/2$.

5 Spectral analysis

Phatarfod and Dyte (1993) determined the eigenvalues, with their multiplicities, for MMTF. We show how to derive these from Theorem 1 and go a step further by explicitly giving the associated idempotents. For $S \subseteq [n]$, we write $\lambda_1(S), \lambda_2(S), \ldots, \lambda_{|S|}(S)$ for the eigenvalues of the principal submatrix R_S of the request matrix R. We will use the notation $1(A)$ for the indicator of A.

Theorem 4 *(a) The set of eigenvalues for the MMTF transition matrix Q defined at (2) is the set of all eigenvalues of all the principal submatrices R_S with $|S| \neq 0, n-1$ of the request chain R. For $S \subseteq [n]$, the eigenvalue $\lambda_i(S)$ corresponding to the ith eigenvalue of R_S has multiplicity in Q equal to $D_{n-|S|}$, the number of derangements (permutations with no fixed points) of $n - |S|$ objects.*

(b) If all the eigenvalues of all of the principal submatrices of R are distinct and nonzero, then MMTF is diagonalizble, with spectral decomposition

$$Q = \sum_{\substack{S \subseteq [n] \\ |S| \neq 0, n-1}} \sum_{i=1}^{|S|} \lambda_i(S) E_{i,S}. \tag{23}$$

Here $E_{i,S}$ is the principal idempotent

$$E_{i,S}(\pi, \sigma) := 1(\sigma[|S|] = S) \sum_{m = |S| \vee L(\pi^{-1}\sigma)}^{n} \sum_{\substack{t_{\to m} \\ t_{|S|} = i}} \frac{H(t_{\to m}, \sigma_{\to m}, \pi_1)}{b_{|S|}(t_{\to m}, \sigma_{\to m})}, \tag{24}$$

where the inner sum in (24) is over all m-tuples $t_{\to m} = (t_1, \ldots, t_m)$ of integers such that $1 \leq t_k \leq k$ for $1 \leq k \leq m$ and $t_{|S|} = i$. Also

$$b_x(t_{\to j}, \sigma_{\to j}) := \prod_{\substack{i \neq x \\ 0 \leq i \leq j}} (\lambda_{t_x}(\sigma[x]) - \lambda_{t_i}(\sigma[i]))$$

with $t_0 := 0$ and $\lambda_0(\emptyset) := 0$, and

$$H(t_{\to j}, \sigma_{\to j}, x) := \left[\prod_{v=1}^{j-1} \sum_{i=1}^{v} R(\sigma_{v+1}, \sigma_i) F_{t_v, \sigma[v]}(\sigma_i, \sigma_v) \right] \sum_{i=1}^{j} R(x, \sigma_i) F_{t_j, \sigma[j]}(\sigma_i, \sigma_j),$$

where $F_{i,S}$ is the principal idempotent of R_S corresponding to $\lambda_i(S)$.

Proof (a) By Theorem 1, for $k \geq 1$,

$$Q^k(\pi,\pi) = \frac{1}{r_{\pi_1}} \sum_{m=1}^{n} P_{\mathbf{r}}[\widetilde{X}_{\widetilde{C}_{\to m}} = \pi_{\to m},\ \widetilde{C}_m \leq k < \widetilde{C}_{m+1},\ \widetilde{X}_k = \pi_1]. \tag{25}$$

Thus

$$\begin{aligned}
\mathrm{tr}(Q^k) &= \sum_{\pi \in S_n} Q^k(\pi,\pi) \\
&= \sum_{m=1}^{n} (n-m)! \sum_{\pi_{\to m}} \frac{1}{r_{\pi_1}} P_{\mathbf{r}}[\widetilde{X}_{\widetilde{C}_{\to m}} = \pi_{\to m},\ \widetilde{C}_m \leq k < \widetilde{C}_{m+1},\ \widetilde{X}_k = \pi_1] \\
&= \sum_{m=1}^{n} (n-m)! \sum_{S \in \binom{[n]}{m}} \sum_{x \in S} P_x[[\widetilde{X}_{\widetilde{C}_{\to m}}] = S,\ \widetilde{C}_m \leq k < \widetilde{C}_{m+1},\ \widetilde{X}_k = x] \\
&= \sum_{m=1}^{n} (n-m)! \sum_{S \in \binom{[n]}{m}} \sum_{x \in S} P_x[[\widetilde{X}_{\to k}] = S,\ \widetilde{X}_k = x],
\end{aligned}$$

where the sum $\sum_{S \in \binom{[n]}{m}}$ is over all m-element subsets S of $[n]$. By inclusion-exclusion, for $x \in S$ we have

$$\begin{aligned}
P_x[[\widetilde{X}_{\to k}] = S,\ \widetilde{X}_k = x] &= \sum_{T \subseteq S} (-1)^{|S|-|T|} P_x[[\widetilde{X}_{\to k}] \subseteq T,\ \widetilde{X}_k = x] \\
&= \sum_{T \subseteq S} (-1)^{|S|-|T|} 1(x \in T) \widetilde{R}_T^k(x,x).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{tr}(Q^k) &= \sum_{m=1}^{n} (n-m)! \sum_{S \in \binom{[n]}{m}} \sum_{T \subseteq S} (-1)^{|S|-|T|} \sum_{x \in T} \widetilde{R}_T^k(x,x) \\
&= \sum_{m=1}^{n} (n-m)! \sum_{S \in \binom{[n]}{m}} \sum_{T \subseteq S} (-1)^{|S|-|T|} \mathrm{tr}(\widetilde{R}_T^k) \\
&= \sum_{\emptyset \neq T \subseteq [n]} \mathrm{tr}(\widetilde{R}_T^k) \sum_{S \supseteq T} (-1)^{|S|-|T|} (n-|S|)! \\
&= \sum_{\emptyset \neq T \subseteq [n]} D_{n-|T|} \mathrm{tr}(\widetilde{R}_T^k).
\end{aligned} \tag{26}$$

It is not hard to check that (26) also holds for $k = 0$. Since $\mathrm{tr}(R_T^k) = \mathrm{tr}(\widetilde{R}_T^k)$, part (a) of Theorem 4 follows.

For part (b) assume that the principal submatrices of R have spectral decompositions given by

$$R_S = \sum_{i=1}^{|S|} \lambda_i(S) F_{i,S}, \quad \text{for } \emptyset \neq S \subseteq [n].$$

Until (28) we use for abbreviation the convention that $\sigma_{m+1} := \pi_1$. From (4) we have

$$\begin{aligned} Q^k(\pi,\sigma;m) &= \sum_{j_{\to m}} \left[\prod_{v=1}^{m} \sum_{i=1}^{v} R(\sigma_{v+1},\sigma_i) R^{j_v}_{\sigma[v]}(\sigma_i,\sigma_v) \right] \\ &= \sum_{j_{\to m}} \left[\prod_{v=1}^{m} \sum_{i=1}^{v} R(\sigma_{v+1},\sigma_i) \sum_{t=1}^{v} \lambda_t^{j_v}(\sigma[v]) F_{t,\sigma[v]}(\sigma_i,\sigma_v) \right] \\ &= \sum_{j_{\to m}} \sum_{t_{\to m}} \left[\prod_{v=1}^{m} \sum_{i=1}^{v} R(\sigma_{v+1},\sigma_i) \lambda_{t_v}^{j_v}(\sigma[v]) F_{t_v,\sigma[v]}(\sigma_i,\sigma_v) \right], \quad (27) \end{aligned}$$

where the second sum is over integer m-tuples $t_{\to m} = (t_1,\ldots,t_m)$ with $1 \le t_v \le v$ for $1 \le v \le m$. Continuing, (27) equals

$$\begin{aligned} &\sum_{t_{\to m}} \left[\sum_{j_{\to m}} \prod_{v=1}^{m} \lambda_{t_v}^{j_v}(\sigma[v]) \right] \prod_{v=1}^{m} \sum_{i=1}^{v} R(\sigma_{v+1},\sigma_i) F_{t_v,\sigma[v]}(\sigma_i,\sigma_v) \\ = &\sum_{t_{\to m}} \sum_{z=0}^{m} \frac{\lambda_{t_z}^k(\sigma[z])}{b_z(t_{\to m},\sigma_{\to m})} H(t_{\to m},\sigma_{\to m},\pi_1), \quad (28) \end{aligned}$$

where for the last equality we have used Proposition A.1 in Fill (1993), with $t_0 := 0, \lambda_0(\emptyset) := 0$, and $b_z(t_{\to m},\sigma_{\to m})$ and $H(t_{\to m},\sigma_{\to m},\pi_1)$ as defined in the statement of the theorem. Thus

$$Q^k(\pi,\sigma) = \sum_{m=L(\pi^{-1}\sigma)}^{n} \sum_{t_{\to m}} \sum_{z=0}^{m} \frac{\lambda_{t_z}^k(\sigma[z])}{b_z(t_{\to m},\sigma_{\to m})} H(t_{\to m},\sigma_{\to m},\pi_1), \quad (29)$$

and by rearrangement and the fact (from (a)) that neither 0 nor any $\lambda_i(S)$ with $|S| = n-1$ is an eigenvalue of Q,

$$\begin{aligned} &Q^k(\pi,\sigma) \\ &= \sum_{t_{\to n}} \sum_{z=0}^{n} \lambda_{t_z}^k(\sigma[z]) \sum_{m=z\vee L(\pi^{-1}\sigma)}^{n} \frac{m! H(t_{\to m},\sigma_{\to m},\pi_1)}{n! b_z(t_{\to m},\sigma_{\to m})} \quad (30) \end{aligned}$$

$$= \sum_{\substack{S\subseteq[n]\\ |S|\neq 0,n-1}} \sum_{i=1}^{|S|} \lambda_i^k(S) 1(\sigma[|S|]=S) \sum_{m=|S|\vee L(\pi^{-1}\sigma)}^{n} \sum_{\substack{t_{\to m}\\ t_{|S|}=i}} \frac{H(t_{\to m},\sigma_{\to m},\pi_1)}{b_{|S|}(t_{\to m},\sigma_{\to m})},$$

where the outer sum in (30) is over all n-tuples $t_{\to n} = (t_1, \ldots, t_n)$ of integers with $1 \le t_z \le z$ for $1 \le z \le n$. ■

Remarks:

1. We stress that Theorem 4(a) makes *no* assumption on the diagonalizability or the eigenvalues of the request chain. The assumption on the eigenvalues in part (b) is for simplicity and convenience. More general cases can be handled with a slight modification of the proof using a perturbation argument.

2. In the i.i.d. case, Theorem 4 recovers results in Fill (1993). In particular, the set of eigenvalues for MTF are all numbers of the form

$$\lambda_S := \sum_{i\in S} r_i$$

with $S \subseteq [n]$ and $|S| \neq n-1$.

6 Separation

6.1 General result

In Sections 6 and 7 we consider two common notions of discrepancy between the distribution of a Markov chain at a fixed time and its stationary distribution. In this subsection we derive an upper bound on separation for MMTF in terms of the separation for the request chain, in the next subsection we apply our result to two examples, and in Section 7 we treat total variation distance. For background on separation see Aldous and Diaconis (1986, 1987). These authors assume ergodicity, which implies that the stationary distribution is unique and strictly positive. We extend the usual definition somewhat.

For a finite-state Markov chain $(Y_n)_{n\ge 0}$ with transition matrix P and unique stationary distribution P^∞, let

$$s_{i,j}(k;P) := \begin{cases} 1 - \frac{P^k(i,j)}{P^\infty(j)}, & \text{if } P^\infty(j) > 0 \\ 0, & \text{otherwise} \end{cases}$$

and define

$$\mathrm{sep}_i(k;P) := \max_j s_{i,j}(k;P)$$

to be the *separation* of the Y-chain at time k when started in state i. We say that a state j^* *achieves* the separation $\mathrm{sep}_i(k;P)$ if $P^\infty(j^*) > 0$ and

$$\mathrm{sep}_i(k;P) = 1 - \frac{P^k(i,j^*)}{P^\infty(j^*)}.$$

We write

$$\mathrm{sep}^*(k;P) := \max_i \mathrm{sep}_i(k;P) = \max_{i,j} s_{i,j}(k;P)$$

for the worst-case separation.

The notion of strong stationary time gives a probabilistic approach to bounding speed of convergence to stationarity for Markov chains. A *strong stationary time* for a Markov chain Y with transition matrix P and unique stationary distribution P^∞, started in state i, is a randomized stopping time T such that Y_T has the distribution P^∞ and is independent of T. Separation can be bounded above by the tail probabilities of a strong stationary time. More precisely,

$$\mathrm{sep}_i(k;P) \le P[T > k] \tag{31}$$

for $k \ge 0$. A fastest strong stationary time, or *time to stationarity*, is a strong stationary time which achieves equality in (31) for all $k \ge 0$. See Diaconis and Fill (1990) for further discussion.

We next review standard terminology in the study of convergence to stationarity. We say that $k = k(n,c)$ steps are sufficient for convergence to stationarity in separation if there exists a function H, independent of n, such that $\mathrm{sep}^*(k;P) \le H(c)$ and $H(c) \to 0$ as $c \to \infty$. We say that $k = k(n,c)$ steps are necessary for convergence to stationarity in separation if there exists a function h, independent of n, such that $\mathrm{sep}^*(k;P) \ge h(c)$ and $h(c) \to 1$ as c tends to $-\infty$ (or whatever the infimum of possible values of c might be.) If $g(n) = o(f(n))$ and $k(n,c) = f(n) + cg(n)$ steps are necessary and sufficient, we say that a "cutoff" occurs at time $f(n)$. Analogous definitions can be given for convergence to stationarity in total variation distance.

The following theorem provides an upper bound on separation for MMTF (Q) in terms of the separation for the request chain (R).

Theorem 5 *Let $\pi \in S_n$. For $\sigma \in S_n$ with $Q^\infty(\sigma) > 0$, let $T(\sigma)$ be a random variable whose distribution is the same as the conditional distribution of $\widetilde{C}_{n-1}$ given $\widetilde{X}_{\widetilde{C}_{\to n}} = \sigma$. Further, for $j \in [n]$ let S_j be a random variable, independent of $T(\sigma)$, whose distribution is that of a fastest strong stationary time for the request chain started in state j. Then*

$$s_{\pi,\sigma}(k; Q) \le P[S_{\pi_1} + T(\sigma) > k],\ k \ge 0. \tag{32}$$

Equality is achieved in (32) for all $k \ge 0$ if and only if $L(\pi^{-1}\sigma) = n-1$ and state σ_{n-1} achieves the separation $sep_{\pi_1}(k; R)$ for every $k > 0$. If these conditions hold and also σ maximizes $\mathcal{L}(T(\sigma))$ stochastically, then

$$\mathrm{sep}_\pi(k; Q) = P[S_{\pi_1} + T(\sigma) > k],\ k \ge 0.$$

Proof Let $\pi, \sigma \in S_n$ and $\rho = r_{\sigma_1}/r_{\pi_1}$. From Theorem 1,

$$\begin{aligned}
Q^k(\pi,\sigma) &\ge \rho \sum_{m=n-1}^{n} P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to m}} = \sigma_{\to m},\, \widetilde{C}_m < k \le \widetilde{C}_{m+1},\, \widetilde{X}_k = \pi_1] \\
&= \rho\{P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to(n-1)}} = \sigma_{\to(n-1)},\, \widetilde{C}_{n-1} < k \le \widetilde{C}_n,\, \widetilde{X}_k = \pi_1] \\
&\quad + P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to n}} = \sigma,\, \widetilde{C}_n < k,\, \widetilde{X}_k = \pi_1]\} \\
&= \rho P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to(n-1)}} = \sigma_{\to(n-1)},\, \widetilde{C}_{n-1} \le k-1,\, \widetilde{X}_k = \pi_1] \\
&= \rho \sum_{t=0}^{k-1} P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to(n-1)}} = \sigma_{\to(n-1)},\, \widetilde{C}_{n-1} = t]\widetilde{R}^{k-t}(\sigma_{n-1}, \pi_1) \\
&= \rho \sum_{t=0}^{k-1} P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to(n-1)}} = \sigma_{\to(n-1)},\, \widetilde{C}_{n-1} = t]\frac{r_{\pi_1}}{r_{\sigma_{n-1}}} R^{k-t}(\pi_1, \sigma_{n-1}) \\
&\ge r_{\sigma_1} \sum_{t=0}^{k-1} P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to(n-1)}} = \sigma_{\to(n-1)},\, \widetilde{C}_{n-1} = t](1 - \mathrm{sep}_{\pi_1}(k-t; R)) \\
&= r_{\sigma_1} P_{\sigma_1}[\widetilde{X}_{\widetilde{C}_{\to n}} = \sigma] \sum_{t=0}^{k-1} P[T(\sigma) = t](1 - \mathrm{sep}_{\pi_1}(k-t; R)) \\
&= Q^\infty(\sigma) \sum_{t=0}^{k} P[T(\sigma) = t](1 - \mathrm{sep}_{\pi_1}(k-t; R)) \\
&= Q^\infty(\sigma) \sum_{t=0}^{k} P[T(\sigma) = t] P[S_{\pi_1} \le k-t],
\end{aligned}$$

where $T(\sigma)$ and S_{π_1} are as defined in the statement of the theorem. The penultimate equality holds since $\mathrm{sep}_{\pi_1}(0;R)=1$, assuming $n\geq 2$. For $k\geq n-1$ the first inequality becomes an equality if and only if $L(\pi^{-1}\sigma)=n-1$ and the second inequality is an equality if and only if state σ_{n-1} achieves the separation $\mathrm{sep}_{\pi_1}(u;R)$ at every strictly positive time u. Since S_{π_1} and $T(\sigma)$ are taken to be independent, the result follows. ■

6.2 Examples

Example 1. I.i.d. case. In the i.i.d. case $S_{\pi_1}\equiv 1$ for all π. Recalling the notation in the first remark of Section 3, the probability generating function for $T(\sigma)$ is given by

$$\sum_{t=0}^{\infty} P[T(\sigma)=t]z^t=\frac{\sum_{t=0}^{\infty} z^t \sum (w_1^+)^{j_1}w_2(w_2^+)^{j_2}w_3\cdots(w_{n-2}^+)^{j_{n-2}}w_{n-1}}{\sum (w_1^+)^{j_1}w_2(w_2^+)^{j_2}w_3\cdots(w_{n-2}^+)^{j_{n-2}}w_{n-1}},$$

where the unmarked sum in the numerator is over all nonnegative $(n-2)$-tuples $j_{\rightarrow(n-2)}$ summing to $t-(n-2)$ and the sum in the denominator is over all nonnegative $(n-2)$-tuples $j_{\rightarrow(n-2)}$. Interchanging sums, the numerator equals the unrestricted sum

$$z^{n-2}\sum (w_1^+z)^{j_1}w_2\cdots(w_{n-2}^+z)^{j_{n-2}}w_{n-1}=z^{n-2}\frac{w_2\cdots w_{n-1}}{(1-w_1^+z)\cdots(1-w_{n-2}^+z)},$$

while the denominator equals

$$\frac{w_2\cdots w_{n-1}}{(1-w_1^+)\cdots(1-w_{n-2}^+)}.$$

Thus

$$\sum_{t=0}^{\infty} P[T(\sigma)=t]z^t=\prod_{v=1}^{n-2}\left[\frac{(1-w_v^+)z}{1-w_v^+z}\right].$$

That is,

$$T(\sigma)\sim \oplus_{v=1}^{n-2}\mathrm{Geom}(1-w_v^+), \tag{33}$$

where the notation indicates the convolution of $n-2$ geometric distributions with the indicated parameters. Hence by Theorem 5,

$$1-\frac{Q^k(\pi,\sigma)}{Q^\infty(\sigma)}\leq P[\oplus_{v=0}^{n-2}\mathrm{Geom}(1-w_v^+)>k], \tag{34}$$

which extends Lemma 4.6 in Fill (1993).

In the i.i.d. case the request chain has separation equal to 0 at all positive times. Thus there is equality in (34) if and only if $L(\pi^{-1}\sigma) = n-1$, from which follows Theorem 4.1 in Fill (1993), which we reproduce for completeness:

Theorem 6 *Consider the move-to-front scheme with weights $r_1, \ldots, r_n$, and suppose (without loss of generality) that $r_1 \geq r_2 \geq \cdots \geq r_n > 0$. Let $\pi \in S_n$ be any permutation with $\pi^{-1}(n-1) > \pi^{-1}(n)$. Then*

$$\mathrm{sep}(k; Q) = 1 - \frac{Q^k(\pi, \mathrm{id})}{Q^\infty(\mathrm{id})} = P[T^* > k], \;\; k = 0, 1, 2, \ldots,$$

where the law of T^* is the convolution of Geometric$(1 - r_v^+)$ distributions, $v = 0, 1, \ldots, n-2$.

Remark:

For general ergodic R we partially generalize the result in (33) by exhibiting the distribution of $T(\sigma)(= \mathcal{L}(\widetilde{C}_{n-1} | X_{\widetilde{C}_{\to n}} = \sigma))$, where $Q^\infty(\sigma) > 0$, as the convolution of $n-2$ distributions. Write $\widetilde{C}_0 := 0$ and let $\widetilde{W}_v := \widetilde{C}_v - \widetilde{C}_{v-1}$ for $1 \leq v \leq n$. Thus $\widetilde{W}_v$ is the waiting time from the $(v-1)$st state covered by $\widetilde{X}$ to the vth, and

$$\widetilde{C}_{n-1} = \sum_{v=1}^{n-1} \widetilde{W}_v = \sum_{v=2}^{n-1} \widetilde{W}_v.$$

But

$$\begin{aligned}
P[\widetilde{W}_2 = w_2, \ldots, \widetilde{W}_{n-1} = w_{n-1} | \widetilde{X}_{\widetilde{C}_{\to n}} = \sigma] & \\
&= \frac{P_{\mathbf{r}}[\widetilde{W}_2 = w_2, \ldots, \widetilde{W}_{n-1} = w_{n-1}, \, \widetilde{X}_{\widetilde{C}_{\to n}} = \sigma]}{P_{\mathbf{r}}[\widetilde{X}_{\widetilde{C}_{\to n}} = \sigma]} \\
&= \frac{w_1 \prod_{v=1}^{n-2} \sum_{i=1}^{v} \widetilde{R}_{\sigma[v]}^{w_{v+1}-1}(\sigma_v, \sigma_i) \widetilde{R}(\sigma_i, \sigma_{v+1})}{w_1 \prod_{v=1}^{n-2} \sum_{i=1}^{v} (I_{\sigma[v]} - \widetilde{R}_{\sigma[v]})^{-1}(\sigma_v, \sigma_i) \widetilde{R}(\sigma_i, \sigma_{v+1})} \\
&= \prod_{v=1}^{n-2} \frac{\sum_{i=1}^{v} \widetilde{R}_{\sigma[v]}^{w_{v+1}-1}(\sigma_v, \sigma_i) \widetilde{R}(\sigma_i, \sigma_{v+1})}{\sum_{i=1}^{v} (I_{\sigma[v]} - \widetilde{R}_{\sigma[v]})^{-1}(\sigma_v, \sigma_i) \widetilde{R}(\sigma_i, \sigma_{v+1})}. \qquad (35)
\end{aligned}$$

It follows that $\widetilde{W}_1, \ldots, \widetilde{W}_{n-1}$ are conditionally independent given $\widetilde{X}_{\widetilde{C}_{\to n}} = \sigma$, with $\widetilde{W}_1 = 0$ and

$$\begin{aligned} P[\widetilde{W}_{v+1} = w | \widetilde{X}_{\widetilde{C}_{\to n}} = \sigma] &= \frac{\sum_{i=1}^{v} \widetilde{R}^{w-1}_{\sigma[v]}(\sigma_v, \sigma_i)\widetilde{R}(\sigma_i, \sigma_{v+1})}{\sum_{i=1}^{v}(I_{\sigma[v]} - \widetilde{R}_{\sigma[v]})^{-1}(\sigma_v, \sigma_i)\widetilde{R}(\sigma_i, \sigma_{v+1})} \quad (36) \\ &= P[\widetilde{W}_{v+1} = w | \widetilde{X}_{\widetilde{C}_{\to (v+1)}} = \sigma_{\to (v+1)}] \end{aligned}$$

for $1 \leq v \leq n-2$.

We remark that in the i.i.d. case, it follows after some calculation from (36) that

$$\mathcal{L}(\widetilde{W}_{v+1} | \widetilde{X}_{\widetilde{C}_{\to n}} = \sigma) = \text{Geom}(1 - w_v^+),$$

as at (33).

Example 2. Random walk. In treating the simple symmetric random walk example of Corollary 4.1, considerable technical difficulties arise involving the endpoints 1 and n of the request chain's state space. We therefore suppose instead that the request chain is simple symmetric *circular* random walk on $[n]$. That is, for fixed $0 < p \leq 1/2$,

$$R(i,j) := \begin{cases} p, & \text{if } j \equiv i-1 \pmod{n} \\ 1-2p, & \text{if } j \equiv i \pmod{n} \\ p, & \text{if } j \equiv i+1 \pmod{n} \\ 0, & \text{otherwise.} \end{cases}$$

An analysis like that in the birth-and-death case shows that $Q^\infty(\sigma) > 0$ if and only if $\sigma_{j+1} \sim \sigma[j]$ for $j \in [n-2]$, in which case

$$Q^\infty(\sigma) = \frac{1}{n!} \prod_{1 \leq j \leq n-2:\ \sigma_j \sim \sigma_{j+1}} j,$$

where now $a \sim b$ means that a is circularly adjacent to b and $a \sim B$ means that $a \sim b$ for some $b \in B$.

Theorem 7 *Consider MMTF whose request chain R is the simple symmetric circular random walk described above. For simplicity assume that $n = 2m$ is even and $0 < p \leq 1/4$.*

(a) A cutoff for separation occurs at time $(18p)^{-1}n^3$*:* $(18p)^{-1}n^3 + cn^{5/2}$ *steps are necessary and sufficient for convergence to stationarity in separation.*

(b) Let $\sigma^* = (2m, 1, 2m-1, 2, \ldots, m+2, m-1, m+1, m)$*. Then the separation* $\mathrm{sep}_{\mathrm{id}}(k; Q)$ *is achieved at* σ^* *for each* $k \geq 0$ *and*

$$\mathrm{sep}^*(k; Q) = \mathrm{sep}_{\mathrm{id}}(k; Q) = P[V > k], \ k \geq 0,$$

where

$$\begin{aligned} \mathcal{L}(V) &= \mathcal{L}(S \oplus T(\sigma^*)) && (37) \\ &= \mathcal{L}(S \oplus W_1 \oplus \cdots \oplus W_{n-2}) && (38) \end{aligned}$$

and the random variables in each sum are mutually independent. The random variable S *is a time to stationarity for the random walk* R *started from any fixed state. The random variable* $T(\sigma^*)$ *is as in Theorem 5. For* $j \in [n-2]$*, the random variable* W_j *has the same distribution as a time to stationarity for the simple symmetric* (non-circular) *random walk of Corollary 4.1, with* n *there replaced by* $v+1$*, started in state 1. Furthermore,*

$$V \sim \oplus_{j=1}^{m} \mathrm{Geom}(1 - \lambda_{j,m}) \oplus \oplus_{l=1}^{n-2} \oplus_{j=1}^{l} \mathrm{Geom}(1 - \lambda_{j,l+1}),$$

where

$$\lambda_{s,t} := 1 - 2p\left(1 - \cos\left(\frac{\pi s}{t}\right)\right).$$

Proof We will first prove (b). After computing the expectation and variance of V, part (a) will follow by Chebychev's inequality.

Suppose that the MMTF chain is started at $\pi = \mathrm{id}$. Noting that $Q^{\infty}(\sigma^*) > 0$, the results of (b) up through (37) will follow from Theorem 5 once we show (i) σ^*_{n-1} achieves the separation $\mathrm{sep}_1(k; R)$ for every $k > 0$; (ii) $L(\sigma^*) = n-1$; and (iii) σ^* maximizes $\mathcal{L}(T(\sigma))$ stochastically. In what follows we use several results from Diaconis and Fill (1990), which treats separation for several classes of Markov chains, including birth-and-death chains. In particular, (i) follows from the discussion in Section 4 of that paper (see especially the end of Example 4.46) where we have used the fact that the holding probability of the random walk satisfies $1/3 \leq 1 - 2p < 1$, a consequence of $0 < p \leq 1/4$, and (ii) is obvious.

To establish (iii) we use the result from the previous remark. Simplifying the notation there, let $W(v,\sigma)$ be a random variable whose distribution is the conditional distribution of $\widetilde{W}_{v+1}$ given $\widetilde{X}_{\widetilde{C}_{\to n}} = \sigma$. It suffices to show that σ^* maximizes $\mathcal{L}(W(v,\sigma))$ stochastically for $v \in [n-2]$ among all $\sigma \in RC = \{\sigma \in S_n : Q^\infty(\sigma) > 0\}$.

For $v \in [n-2]$ and $\sigma \in RC$, $R_{\sigma[v]}$ does not depend on σ. Write R_v for such $R_{\sigma[v]}$, and relabel the rows and columns sequentially with $1, \ldots, v$. Thus R_v is a tridiagonal $v \times v$ matrix with diagonal entries $1-2p$ and sub- and super-diagonal entries p. Arguing as in Feller (1968, Section XVI.3), we diagonalize R_v (omitting details) to obtain, for $k \geq 1$ and $i, j \in [v]$,

$$R_v^k(i,j) = \frac{2}{v+1} \sum_{l=1}^{v} \sin\left(\frac{\pi l i}{v+1}\right) \sin\left(\frac{\pi l j}{v+1}\right) \lambda_{l,v+1}^k.$$

Since $|\lambda_{l,v+1}| < 1$ for all $l \in [v]$,

$$\begin{aligned}(I_v - R_v)^{-1}(i,j) &= \sum_{k=0}^{\infty} R_v^k(i,j) \\ &= \frac{1}{p(v+1)} \sum_{l=1}^{v} \sin\left(\frac{\pi l i}{v+1}\right) \sin\left(\frac{\pi l j}{v+1}\right) \left(1 - \cos\left(\frac{\pi l}{v+1}\right)\right)^{-1},\end{aligned}$$

and hence all the ingredients of (36) can be calculated explicitly. For fixed v and $\sigma \in RC$,

$$P[W(v,\sigma) = w] = \begin{cases} \frac{R_v^{w-1}(1,1)}{(I_v - R_v)^{-1}(1,1)}, & \text{if } \sigma_{v+1} \sim \sigma_v \\ \frac{R_v^{w-1}(1,v)}{(I_v - R_v)^{-1}(1,v)}, & \text{if } \sigma_{v+1} \not\sim \sigma_v. \end{cases}$$

It is obvious from a directly probabilistic argument that

$$\mathcal{L}(W(v,\sigma)) = \mathcal{L}(\widetilde{W}_{v+1} | \widetilde{X}_{\widetilde{C}_{\to(v+1)}})$$

is stochastically larger in the case $\sigma_{v+1} \not\sim \sigma_v$ than in the case $\sigma_{v+1} \sim \sigma_v$. Since $\sigma^*_{v+1} \not\sim \sigma^*_v$ for all $v \in [n-2]$, condition (iii) is met.

It is now straightforward to compute

$$\begin{aligned}P[W(v,\sigma^*) = w] &= 2p \sum_{l=1}^{v} (-1)^{l-1} \sin^2\left(\frac{\pi l}{v+1}\right) \lambda_{l,v+1}^{w-1} \\ &= \sum_{l=1}^{v} \alpha(l)(1 - \lambda_{l,v+1}) \lambda_{l,v+1}^{w-1}, \qquad (39)\end{aligned}$$

where

$$\alpha(l) := (-1)^{l-1}\left(1 + \cos\left(\frac{\pi l}{v+1}\right)\right), \quad l \in [v].$$

This distribution has a curious and useful interpretation. As shown in Section 4 of Diaconis and Fill (1990), if the *noncircular* walk (call it R'_{v+1}) described in Corollary 4.1, with n there replaced by $v+1$, is started in state 1, then for every $k \geq 0$ the separation $\mathrm{sep}_1(k; R'_{v+1})$ is achieved at state $v+1$ and so equals $1 - (v+1)(R'_{v+1})^k(1, v+1)$. Diagonalizing R'_{v+1}, one discovers

$$\mathrm{sep}_1(k; R'_{v+1}) = P[W(v, \sigma^*) > k], \quad k \geq 0,$$

using (39) for the right side. Thus $W(v, \sigma^*)$ is distributed as the time to stationarity for R'_{v+1} started at 1.

By (4.58) in Diaconis and Fill (1990) (in which the sum should be $\sum_{j=1}^{d}$) and the ensuing discussion, for every $k \geq 0$ the separation $\mathrm{sep}_1(k; R)$ equals

$$\mathrm{sep}_1(k; R) = 2\sum_{j=1}^{m}(-1)^{j-1}\cos\left(\frac{j\pi}{2m}\right)\left(1 - 2p\left(1 - \cos\left(\frac{j\pi}{m}\right)\right)\right)^k$$

and is achieved (uniquely) at state $m+1$.

Again by the discussion in Diaconis and Fill (1990) (see especially Theorem 4.20), the probability generating function for S is given by

$$z \mapsto \prod_{j=1}^{m} \frac{(1 - \lambda_{j,m})z}{1 - \lambda_{j,m}z},$$

and W_l has probability generating function

$$z \mapsto \prod_{j=1}^{l} \frac{(1 - \lambda_{j,l+1})z}{1 - \lambda_{j,l+1}z}.$$

Since we assume $p \leq 1/4$, $\lambda_{s,t} \geq 0$ for all s and t and thus S and each W_l is distributed as the sum of independent geometric random variables, completing the proof of (b). When some of the λ's are negative, a distributional interpretation can be provided along the lines of Remark 4.22(c) in Diaconis and Fill (1990).

For (a) we note that $k(n,c) = E[V] + c\sqrt{\mathrm{Var}[V]}$ steps are necessary and sufficient by Chebychev's inequality. But

$$\begin{aligned} E[V] &= \sum_{j=1}^{m} \frac{1}{1-\lambda_{j,m}} + \sum_{l=1}^{n-2} \sum_{j=1}^{l} \frac{1}{1-\lambda_{j,l+1}} \\ &= \frac{1}{2p} \sum_{j=1}^{n/2} \frac{1}{1-\cos(2\pi j/n)} + \frac{1}{2p} \sum_{l=1}^{n-2} \sum_{j=1}^{l} \frac{1}{1-\cos(\pi j/(l+1))} \\ &= \frac{n^3}{18p} + O(n^2) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}[V] &= \sum_{j=1}^{m} \frac{\lambda_{j,m}}{(1-\lambda_{j,m})^2} + \sum_{l=1}^{n-2} \sum_{j=1}^{l} \frac{\lambda_{j,l+1}}{(1-\lambda_{j,l+1})^2} \\ &= \frac{1}{4p^2} \sum_{j=1}^{n/2} \frac{1-2p(1-\cos(2\pi j/n))}{(1-\cos(2\pi j/n))^2} + \frac{1}{4p^2} \sum_{l=1}^{n-2} \sum_{j=1}^{l} \frac{1-2p(1-\cos(\pi j/(l+1)))}{(1-\cos(\pi j/(l+1)))^2} \\ &= \frac{n^5}{450p^2} + O(n^4), \end{aligned}$$

using standard asymptotic analysis and the values $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$ and $\sum_{k=1}^{\infty} k^{-4} = \pi^4/90$ of the Riemann zeta function. ■

Remarks:

1. An alternative approach applying Markov's inequality to

$$\exp\left[\text{const.} \times \frac{V - E[V]}{\sqrt{\mathrm{Var}[V]}}\right]$$

shows that $H(c)$ in the definition of "sufficient number of steps" can be taken to be of the form $H(c) = \alpha e^{-\beta c}$ where α and β are positive constants. Such exponential decay of H is sometimes required in the definition of "sufficient number of steps."

2. An even sharper result, at least asymptotically as $n \to \infty$, is that for each fixed $c \in \mathbb{R}$ and $k = \lfloor (18p)^{-1} n^3 + c(15\sqrt{2}\ p)^{-1} n^{5/2} \rfloor$,

$$\mathrm{sep}^*(k; Q) \to P[Z > c] \text{ as } n \to \infty,$$

where Z is a standard normal random variable. This can be shown using Liapounov's doubly indexed array version of the central limit theorem (e.g., Chung (1974, Section 7.2)) and a calculation that the sum of the fourth central moments of the component geometric random variables is of order n^9.

7 Total variation distance

In this section we use coupling to derive bounds on the total variation distance between MMTF and its stationary distribution.

Let $X = (X_n)$ and $Y = (Y_n)$ be two realizations of a Markov chain with transition matrix P. Suppose that the X-chain has an arbitrary initial distribution and the Y-chain is started (say) in stationarity. A *coupling time* T is a (randomized) stopping time for the bivariate chain (X_n, Y_n) such that $X_n = Y_n$ for all $n \geq T$.

Our coupling for the MMTF chain is constructed in two steps: (i) First couple two copies of the request chain; (ii) then use the standard coupling for MTF, namely, wait until all but one of the records have been requested at least once. Thus the analysis for the MMTF chain reduces to an analysis of the much smaller request chain.

Let P^∞ denote the stationary distribution of the chain and let $P_i^k = P^k(i, \cdot)$ be the distribution after k steps when the chain is started in state i. The *total variation distance* between P_i^k and P^∞ is

$$d_i(k; P) := \|P_i^k - P^\infty\|_{TV} = \max_A |P_i^k(A) - P^\infty(A)| = \min_T P[T > k], \quad (40)$$

where the maximum in the third expression is over all events A and the minimum in the fourth expression is over all coupling times T for P_i^k and P^∞.

For the request chain X run in stationarity and $t \in \{0, 1, 2, \ldots\}$, let

$$C(t) := \inf\{k \geq 0 : \{X_t, X_{t+1}, \ldots, X_{t+k}\} = [n]\}$$

be the *ergodic cover time* for the chain X started at time t. Let $C := C(0)$. The following result is evident.

Theorem 8 *For $i \in [n]$, let F_i be a fastest coupling time for the request chain started at state i. For $t \in \{0, 1, 2, \ldots\}$, let $C(t)$ be the ergodic cover time of the request chain started at time t. Let $\pi \in S_n$. Then $T_{\pi_1} := F_{\pi_1} + C(F_{\pi_1})$ is a coupling time for MMTF started in π. Furthermore, for any $\zeta \in [0, 1)$,*

$$\begin{aligned} d_\pi(k; Q) &\leq P[T_{\pi_1} > k] \\ &\leq P[F_{\pi_1} > \zeta k] + P[C > (1-\zeta)k] \\ &\leq d_{\pi_1}(\lfloor \zeta k \rfloor; R) + \frac{1}{(1-\zeta)k} E[C]. \end{aligned} \tag{41}$$

For our running random walk example, direct arguments give the following.

Theorem 9 *Consider MMTF with request chain R as given in Corollary 4.1. Then cn^2 steps are necessary and sufficient for convergence to stationarity in total variation distance.*

Proof The proof we sketch is based on Theorem 8 and well-known results for simple symmetric random walk on $[n]$.

For the lower bound, it takes cn^2 steps for the front record in the list, from any fixed initial state, to become uniform in total variation distance. Thus it takes at least cn^2 steps for the list as a whole to become stationary.

For the upper bound, cn^2 steps are sufficient for the request chain to become uniform in total variation distance. So, from any initial i, the fastest coupling F_i takes at most cn^2 steps. And cn^2 steps are always sufficient to cover $[n]$. The upper bound thus follows from Theorem 8. ■

Remark:

It is interesting to note the different behaviors for separation and total variation distance for this example. Convergence to stationarity in total variation distance exhibits no cutoff. On the other hand, convergence to stationarity in separation exhibits a cutoff and is slower by an order of magnitude.

Application of Theorem 8 (e.g., using the second and third inequalities in (41)) to more general request chains requires knowledge of the covering time C. Unfortunately, there are few results on the distribution of C for any but

the most structured chains. There is a growing body of work which treats expected cover time, but these bounds are typically not sharp. A valuable source for such results is Broder and Karlin (1989).

Without going into details, this approach gives an upper bound in the case of the mixture model (1) introduced in Section 2 when R_0 is the matrix each of whose entries is $1/n$ and B is the request transition matrix discussed in Corollary 4.1. For fixed $\alpha \in (0,1)$, order $n \log n$ steps are sufficient for convergence to stationarity in total variation distance. We conjecture that order $n \log n$ steps are also necessary.

8 References

Aldous, D., and Diaconis, P. (1986). Shuffling cards and stopping times. *Amer. Math. Monthly* **93** 333–347.

Aldous, D., and Diaconis, P. (1987). Strong uniform times and finite random walks. *Adv. in Appl. Math.* **8** 69–97.

Bentley, J. L., and McGeoch, C. C. (1985). Amortized analyses of self-organizing sequential search heuristics. *Comm. ACM* **29** 404–411.

Bitner, J. R. (1979). Heuristcs that dynamically organize data structures. *SIAM J. Comp.* **8** 82–110.

Broder, A. Z., and Karlin, A. R. (1989). Bounds on the cover time. *J. Theo. Prob.* **2** 101–120.

Chung, K. L. (1974). *A Course in Probability Theory,* 2nd edition. Academic Press, New York.

Diaconis, P. (1993). Notes on the weighted rearrangement process. Unpublished manuscript.

Diaconis, P., and Fill, J. A. (1990). Strong stationary times via a new form of duality. *Ann. Prob.* **18** 1483–1522.

Diaconis, P., Fill, J. A., and Pitman, J. (1992). Analysis of top to random shuffles. *Comb., Prob. and Comp.* **1** 135–155.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications,* Vol. I. John Wiley & Sons, New York.

Fill, J. A. (1993). An exact formula for the move-to-front rule for self-organizing lists. Technical Report #529, Department of Mathematical Sciences, The Johns Hopkins University.

Hendricks, W. J. (1989). *Self-organizing Markov Chains.* MITRE Corp., McLean, Va.

Hester, J. H., and Hirschberg, D. S. (1985). Self-organizing linear search. *Comp. Surveys* **17** 295–311.

Kapoor, S., and Reingold, E. M. (1991). Stochastic rearrangement rules for self-organizing data structures. *Algorithmica* **6** 278–291.

Knuth, D. (1973). *The Art of Computer Programming, Searching and Sorting,* Vol. III. Addison-Wesley, Reading, Mass.

Konnecker, L. K., and Varol, Y. L. (1981). A note on heuristics for dynamic organization of data structures. *Info. Proc. Lett.* **12** 213–216.

Lam, K., Leung, M.-Y., and Siu, M.-K. (1984). Self-organizing files with dependent accesses. *J. Appl. Prob.* **21** 343–359.

Phatarfod, R. M., and Dyte, D. (1993). The linear search problem with Markov dependent requests. Preprint.

Rivest, R. (1978). On self-organizing sequential search heuristics. *Comm. ACM* **19** 63–67.

Robert P. Dobrow
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218-2689

James Allen Fill
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218-2689