# Exit frequency matrices for finite Markov chains

Andrew Beveridge[*]
Department of Mathematics and Computer Science
Macalester College
Saint Paul, MN 55105

László Lovász[†]
Institute of Mathematics
Eötvös Loránd University
Budapest, Hungary

**Abstract**

Consider a finite irreducible Markov chain on state space $S$ with transition matrix $M$ and stationary distribution $\pi$. Let $R$ be the diagonal matrix of return times, $R_{ii} = 1/\pi_i$. Given distributions $\sigma, \tau$ and $k \in S$, the exit frequency $x_k(\sigma, \tau)$ denotes the expected number of times a random walk exits state $k$ before an optimal stopping rule from $\sigma$ to $\tau$ halts the walk. For a target distribution $\tau$, we define $X_\tau$ as the $n \times n$ matrix given by $(X_\tau)_{ij} = x_j(i, \tau)$, where $i$ also denotes the singleton distribution on state $i$.

The dual Markov chain with transition matrix $\hat{M} = RM^\top R^{-1}$ is called the *reverse chain*. We prove that Markov chain duality extends to matrices of exit frequencies. Specifically, for each target distribution $\tau$, we associate a unique dual distribution $\tau^*$. Let $\hat{X}_{\tau^*}$ denote the matrix of exit frequencies from singletons to $\tau^*$ on the reverse chain. We show that $\hat{X}_{\tau^*} = R(X_\tau^\top - \mathbf{b}^\top \mathbf{1})R^{-1}$ where $\mathbf{b}$ is a nonnegative constant vector (depending on $\tau$). We explore this exit frequency duality and further illuminate the relationship between stopping rules on the original chain and reverse chain.

# Contents

# 1 Stopping rules for finite Markov chains

## 1.1 Introduction

Consider a finite, irreducible, discrete time Markov chain on the state space $S$ where $|S| = n$ with transition matrix $M = (p_{ij})$ and stationary distribution $\pi$. Reversing time, we get the dual Markov chain, called the reverse chain, with transition matrix $\hat{M} = (\hat{p}_{ij})$ where $\hat{p}_{ij} = \pi_j p_{ji}/\pi_i$. $\hat{M}$ has the same stationary distribution as the original chain: $\hat{\pi} = \pi$. In what follows, hatted symbols always refer to the reverse chain. Note that

$$\hat{M} = RM^\top R^{-1} \tag{1}$$

where $R$ is the diagonal matrix of return times, $R_{ii} = \mathrm{Ret}(i) = 1/\pi_i$.

In [12], Lovász and Winkler consider the relationship between time reversal and exact mixing measures defined via stopping rules (stopping times). Given a starting distribution $\sigma$ and a target distribution $\tau$, a *stopping rule* halts the Markov chain whose initial state is drawn from $\sigma$ so that the final state is governed by $\tau$ (see [10, 11] for details). An *optimal* stopping rule from $\sigma$ to $\tau$ minimizes the expected time before the rule halts. The *access time* is the expected length of such an optimal stopping rule, which we denote by $H(\sigma, \tau)$. We may think of the access time as a generalization of the state-to-state hitting time $H(i, j)$.

Considering a target distribution that captures an aspect of mixing leads to a number of parameterless mixing measures. Three of the most important measures are the *mixing time*

$$T_{\mathrm{mix}} = \max_{i \in S} H(i, \pi),$$

the *reset time*

$$T_{\mathrm{reset}} = \sum_{j \in S} \pi_j H(j, \pi)$$

and the *forget time*

$$T_{\mathrm{forget}} = \min_{\tau} \max_{i \in S} H(i, \tau)$$

where we minimize $\tau$ over all distributions on $S$. We interpret $T_{\mathrm{mix}}$ as the pessimal mixing time and $T_{\mathrm{reset}}$ as the average mixing time. The forget time $T_{\mathrm{forget}}$ is the minimum expected time to achieve some distribution regardless of our initial state (thus "forgetting" our starting point). Lovász and Winkler [12] prove the following two results.

**Theorem 1.1** *For every finite Markov chain, $T_{\mathrm{mix}} = \hat{T}_{\mathrm{mix}}$.*

**Theorem 1.2** *For every finite Markov chain, $T_{\mathrm{forget}} = \hat{T}_{\mathrm{reset}}$ and $T_{\mathrm{reset}} = \hat{T}_{\mathrm{forget}}$. Moreover, $T_{\mathrm{forget}}$ is achieved uniquely by the target distribution $\mu$ given by*

$$\mu_i = \pi_i \left( 1 + \sum_{j \in S} \hat{p}_{ij} \hat{H}(j, \pi) - \hat{H}(i, \pi) \right) \tag{2}$$

*By Markov chain duality, $\hat{T}_{\mathrm{forget}}$ is achieved uniquely by the target distribution $\hat{\mu}$ given by*

$$\hat{\mu}_i = \pi_i \left( 1 + \sum_{j \in S} p_{ij} H(j, \pi) - H(i, \pi) \right). \tag{3}$$

We interpret $\hat{\mu}_i$ as the scaled difference between two rules from $i$ to $\pi$ on the forward chain. $H(i, \pi)$ is the expected length of an optimal rule from $i$ to $\pi$, while the quantity $1 + \sum_{j \in S} p_{ij} H(j, \pi)$ is the expected length of the rule "make one transition according to $M$ and then follow an optimal rule from this state to $\pi$." Thus the quantity $\hat{\mu}_i / \pi_i$ measures the distance from optimality of the latter rule. We can think of $\mu_i$ as contrasting the expected length of an optimal rule with the expected length of a second (and possibly non-optimal) rule from $i$ to $\pi$.

The goal of this paper is to put constructions (2) and (3) in a general framework. The original proof of Theorem 1.2 relies on a difficult linear programming duality argument. The general framework developed herein provides a new, simple proof of Theorem 1.2 using Markov chain duality. We also provide a new proof of Theorem 1.1. Our main tool for expressing this duality is the notion of *optimal exit frequency matrices* (defined in the next section). Quite succinctly, we show that the duality of the pair $\pi, \hat{\mu}$ expressed in Theorem 1.2 is analogous to the duality of equation (1). Furthermore, this duality can be extended to other pairs of distributions.

The organization of this paper is the following. In Section 1.2, we recall some fundamental results about exit frequencies and introduce the key concept of exit frequency matrices. Section 1.3 enumerates some known results concerning hitting times and access times.

Section 2 lays out the general framework of our results. In Section 2.1, we introduce the *contrast map*, which generalizes the construction of Theorem 1.2. Section 2.2 develops the fundamental duality results of our general framework. We introduce *contrasting pairs* of distributions and prove that their optimal exit frequency matrices are related by scaled transposition. This result parallels the relationship between the transition matrices for the forward and reverse chains. Section 2.3 contains several examples of contrasting pairs, and Section 2.4 further develops their duality.

Section 3 contains the new proofs of the mixing results of Theorems 1.1 and 1.2. Finally, Section 4 studies the matrix structure of exit frequency matrices.

## 1.2  Exit frequency matrices

We begin by briefly summarizing some essential results from [11]. Given a $(\sigma, \tau)$-stopping rule $\Gamma$, we partition its expected length according to the expected number of exits at each state. For

each $i \in S$ we define its $i$th *exit frequency* $x_i(\Gamma)$ to be the expected number of times the walk leaves state $i$ before halting. A key observation, due to Pitman [14], is that exit frequencies satisfy the *conservation equation*

$$\sum_{i \in S} p_{ij} x_i(\Gamma) - x_j(\Gamma) = \tau_j - \sigma_j \tag{4}$$

for every $j \in S$. Moreover, [10] describes multiple ways to construct stopping rules from a given set of desired exit frequencies. That is, if we have a set of nonnegative numbers $\{x_i\}_{i \in S}$ such that $\sum_{i \in S} p_{ij} x_i - x_j = \tau_j - \sigma_j$ for all $j \in S$, then we can construct a $(\sigma, \tau)$-stopping rule $\Gamma$ such that $x_i(\Gamma) = x_i$ for all $i \in S$.

The conservation equation is a very useful tool. For example, let $\Gamma$ and $\Gamma'$ be two $(\sigma, \tau)$-stopping rules with respective exit frequencies $(x_i)_{i \in S}$ and $(x_i')_{i \in S}$, expressed as vectors. The conservation equation guarantees that

$$(x_i)_{i \in S} - (x_i')_{i \in S} = K\pi \text{ where } K = E(\Gamma) - E(\Gamma'). \tag{5}$$

Subsequently, we conclude that every optimal $(\sigma, \tau)$ stopping rule has the same exit frequencies. We denote these optimal exit frequencies by $x_i(\sigma, \tau)$. Moreover,

$$\Gamma \text{ is an optimal stopping rule} \iff \exists\, k \in S,\; x_k(\Gamma) = 0. \tag{6}$$

When $x_k(\Gamma) = 0$, we call the state $k$ a $(\sigma, \tau)$-*halting state*, or simply a *halting state* when the initial and target distributions are clear. The presence of a halting state is the single most useful criterion for determining whether a given rule is optimal. The following formula (Theorem 4.4 in [11]) expresses exit frequencies in terms of access times:

$$x_k(\sigma, \tau) = \pi_k(H(\sigma, \tau) + H(\tau, k) - H(\sigma, k)). \tag{7}$$

Sometimes it is more convenient to consider the *scaled exit frequencies*

$$y_k(\sigma, \tau) = \frac{1}{\pi_k} x_k(\sigma, \tau). \tag{8}$$

In the remainder of this section, we introduce *exit frequency matrices* and develop some basic properties. Fixing the target distribution $\tau$, our key idea is to consider a family of rules from singletons to $\tau$ as an ensemble. This novel organization of a family of rules is the key to unlocking the duality between exit frequencies of stopping rules.

5

For convenience, we use the symbol $i$ to denote the state $i \in S$ as well as the singleton distribution concentrated at state $i$. This minor abuse of notation leads to clearer formulations of our results.

**Definition 1.3** A *(forward) $\tau$-family* is a set of stopping rules $\Gamma(S, \tau) = \{\Gamma(i, \tau)\}_{i \in S}$, one from each singleton to $\tau$. The *(forward) exit frequency matrix* $X_{\Gamma(S,\tau)}$ is the $n \times n$ matrix with

$$\left(X_{\Gamma(S,\tau)}\right)_{ij} = x_j(\Gamma(i, \tau)) \text{ for all } i, j \in S.$$

An *optimal $\tau$-family* is a $\tau$-family in which every $\Gamma(i, \tau)$ is an optimal stopping rule. Its corresponding *optimal exit frequency matrix* is denoted by $X_\tau$, so that

$$(X_\tau)_{ij} = x_j(i, \tau) \text{ for all } i, j \in S.$$

We make three observations. First, the exit frequency matrix organizes the rules by rows: the $i$th row of $X_{\Gamma(S,\tau)}$ contains the exit frequencies for $\Gamma(i, \tau)$. Second, each row of the optimal exit frequency matrix $X_\tau$ must contain at least one zero entry, corresponding to a halting state for this optimal rule. Third, the analogous definitions hold for a *reverse (optimal) $\tau$-family*, and a *reverse (optimal) exit frequency matrix*, associated with the the reverse chain. Since our designation "forward" and "reverse" is arbitrary, most of the discussion hereafter will assume that we are considering stopping rules for $\tau$ on the forward chain. The analogous statements concerning the reverse chain also hold, but we typically omit their formulation.

For any $\tau$-family, we may rewrite the conservation equation (4) in matrix form:

$$X_{\Gamma(S,\tau)}(I - M) = I - \mathbf{1}\,\tau^\top. \tag{9}$$

We can immediately use this matrix formulation to prove a useful lemma.

**Lemma 1.4** *For any distribution $\sigma$ on the forward chain $M$ and any distribution $\hat{\tau}$ on the reverse chain $\hat{M}$,*

$$(I - \mathbf{1}\sigma^\top)R\hat{X}_{\hat{\tau}}^\top = X_\sigma R(I - \hat{\tau}\mathbf{1}^\top).$$

**Proof.** Transpose the matrix conservation equation (9) for the reverse walk to $\hat{\tau}$ and substitute $\hat{M}^\top = R^{-1}MR$ to yield $R^{-1}(I - M)R\hat{X}_{\hat{\tau}}^\top = I - \hat{\tau}\mathbf{1}^\top$. Multiplying on the left by $X_\sigma R$ and using equation (9) for $\sigma$ gives the result. □

6

The $ij$th entry gives a special case of the "duality formula" in Lemma 4.1 of [12]:

$$\hat{y}_i(j, \hat{\tau}) - \sum_{k \in S} \sigma_k \hat{y}_k(j, \hat{\tau}) = y_j(i, \sigma) - \sum_{k \in S} \hat{\tau}_k y_k(i, \sigma). \tag{10}$$

## 1.3 Hitting times and access times

We collect some additional results about hitting times and access times that will be referenced herein, and prove one new result about *naive access times*. For an introduction to finite Markov chains, see [1, 7, 9]. For a thorough introduction to stopping rules, see [11].

The sum $\kappa(i, j) = H(i, j) + H(j, i)$ is the *commute time*. As an example of the duality of the forward and reverse chains, we have

$$\kappa(i, j) = \hat{\kappa}(i, j) \tag{11}$$

for all $i, j \in S$. We also have the *cycle reversing identity* [5]:

$$H(i, j) + H(j, k) + H(k, i) = \hat{H}(i, k) + \hat{H}(k, j) + \hat{H}(j, i) \tag{12}$$

for all states $i, j, k \in S$. The analogous identity holds for more than three states (with a proof that is virtually identical to proof of equation (12) found in [5]). Another useful formula is the *random target identity* (cf. [1], Chapter 2, Corollary 14) which states that the quantity

$$T_{\text{hit}} = \sum_{j \in S} \pi_j H(i, j) \tag{13}$$

is independent of the starting state $i$. Moreover,

$$T_{\text{hit}} = \hat{T}_{\text{hit}}. \tag{14}$$

We now use our version of the duality formula (10) to prove a generalization of equations (11) and (14). Given distributions $\sigma$ and $\tau$, the *naive rule* from $\sigma$ to $\tau$ is "choose a target state $j$ according to $\tau$ and then walk until you reach $j$." The expected length of such a rule is the *naive access time*

$$N(\sigma, \tau) := \sum_{i \in S} \sum_{j \in S} \sigma_i \tau_j H(i, j). \tag{15}$$

**Proposition 1.5** *For any distribution $\tau$, we have $N(\tau, \tau) = \hat{N}(\tau, \tau)$.*

**Proof.** Consider the duality formula (10) with $i = j$ and $\sigma = \hat{\tau} = \tau$. Weight this equation by $\tau_i$ and sum over $i \in S$ to get

$$\sum_{i \in S} \tau_i \hat{y}_i(i, \tau) - \sum_{i \in S} \sum_{k \in S} \tau_i \tau_k \hat{y}_k(i, \tau) = \sum_{i \in S} \tau_i y_i(i, \tau) - \sum_{i \in S} \sum_{k \in S} \tau_i \tau_k y_k(i, \tau)$$

$$\sum_{i \in S} \sum_{k \in S} \tau_i \tau_k \hat{H}(i, k) = \sum_{i \in S} \sum_{k \in S} \tau_i \tau_k H(i, k),$$

where we use equation (7) to expand the scaled exit frequencies. $\square$

Taking $\tau_i = \tau_j = 1/2$ (and 0 elsewhere), we recover equation (10). Taking $\tau = \pi$, we recover equation (14).

The following results can be found in [11]. These results are essential for finding optimal stopping rules and calculating their access times. These facts will be used frequently hereafter. We have the following formula for access times:

$$H(\sigma, \tau) = \max_{j \in S} (H(\sigma, j) - H(\tau, j)). \tag{16}$$

When $\sigma = i$ and $\tau = \pi$, we have a nice characterization of the halting states:

$$j \text{ is an } (i, \pi)\text{-halting state} \iff H(j, i) = \max_{k \in S} H(k, i). \tag{17}$$

We use $i'$ to denote an *i-pessimal node* achieving

$$H(i', i) = \max_{k \in S} H(k, i). \tag{18}$$

We have a very useful formula for the access time from a singleton to $\pi$:

$$H(i, \pi) = H(i', i) - H(\pi, i). \tag{19}$$

## 2 Exit frequency matrices and contrasting pairs

### 2.1 The contrast map

In this section, we introduce the *contrast map*, which associates a unique distribution $\tau^*$ to $\tau$.

We extend the "hatted" notation of Theorem 1.1 and Theorem 1.2 to distinguish between distributions on the forward chain and distributions on the reverse chain. That is, the symbol $\tau$ will represent a distribution on the state space of the forward chain, while $\hat{\tau}$ denotes a distribution on the state space of the reverse chain. This usage is consistent with the definitions of $\mu$ and $\hat{\mu}$ above.

The following two maps are inspired by equations (2) and (3).

**Definition 2.1** Let $\tau$ be a distribution on the forward chain. The *forward contrast map* is the function $c^* : \tau \mapsto \tau^*$ where

$$\tau_i^* = \pi_i \left( 1 + \sum_{j \in S} p_{ij} H(j, \tau) - H(i, \tau) \right). \tag{20}$$

We call $\tau^*$ the *forward contrast distribution of $\tau$* or the *forward $\tau$-contrast.*

Let $\hat{\tau}$ be a distribution of the reverse chain. The *reverse contrast map* is the function $\hat{c}^* : \hat{\tau} \mapsto \hat{\tau}^*$ where

$$\hat{\tau}_i^* = \pi_i \left( 1 + \sum_{j \in S} \hat{p}_{ij} \hat{H}(j, \hat{\tau}) - \hat{H}(i, \hat{\tau}) \right). \tag{21}$$

We call $\hat{\tau}^*$ the *reverse contrast distribution of $\hat{\tau}$* or the *reverse $\hat{\tau}$-contrast.*

Note that the forget distribution equations (2) and (3) are a special case of the contrast map. Indeed, we have $\mu = \hat{c}^*(\hat{\pi}) = \hat{\pi}^*$ and $\hat{\mu} = c^*(\pi) = \pi^*$.

We are guaranteed that $\tau^*$ and $\hat{\tau}^*$ are distributions. Indeed, the value of $\tau_i^*$ is the scaled difference between the expected length of the (possibly non-optimal) rule "make one transition according to $M$ and then follow an optimal rule from this new state to $\tau$" and the expected length of an optimal $(i, \tau)$-rule, which makes it nonnegative. Since $\sum_{i \in S} p_{ij} \pi_i = \pi_j$ for every $j$ and $\pi^\top M = \pi^\top$, we have

$$\begin{aligned}
\sum_{i \in S} \tau_i^* &= \sum_{i \in S} \pi_i \left( 1 + \sum_{j \in S} p_{ij} H(j, \tau) - H(i, \tau) \right) \\
&= \sum_{i \in S} \pi_i + \sum_{i \in S} \sum_{j \in S} p_{ij} \pi_i H(j, \tau) - \sum_{i \in S} \pi_i H(i, \tau) = 1.
\end{aligned}$$

The argument for $\hat{\tau}^*$ is similar.

We follow up our definitions of the contrast maps with a straightforward generalization of Theorem 7.1 in [12]. This result suggests that $\tau^*$ belongs on the reverse chain.

**Proposition 2.2** *The exit frequencies for the reverse walk from $\pi$ to $\tau^*$ and vice versa are given by*

$$\hat{x}_i(\pi, \tau^*) = \pi_i \left( H(i, \tau) - \min_{k \in S} H(k, \tau) \right)$$

*and*

$$\hat{x}_i(\tau^*, \pi) = \pi_i \left( \max_{k \in S} H(k, \tau) - H(i, \tau) \right).$$

**Proof.** Rewriting equation (20) gives

$$\sum_{j \in S} \hat{p}_{ji} \pi_j H(j, \tau) - \pi_i H(i, \tau) = \tau_i^* - \pi_i.$$

The set $\{\hat{x}_i\}_{i \in S} := \{\pi_i H(i, \tau)\}_{i \in S}$ satisfies the $(\pi, \tau^*)$ reverse conservation equation (defined analogously to equation (4)). As shown in [10], we can construct a reverse $(\pi, \tau^*)$-stopping rule with the exit frequencies $\{\hat{x}_i\}_{i \in S}$. Using equation (5), we conclude that $\pi_i \left( H(i, \tau) - \min_{k \in S} H(k, \tau) \right)$ is the $i$th reverse exit frequency for an optimal stopping rule.

Taking the negative of both sides and following a similar argument gives the second statement of the Theorem. □

We close this subsection with one final collection of definitions.

**Definition 2.3** A distribution $\hat{\rho}$ on the reverse chain is *forward contrasted* if there exists some distribution $\tau$ (on the forward chain) such that $\hat{\rho} = \tau^*$. A distribution $\rho$ on the forward chain is *reverse contrasted* if there exists some distribution $\hat{\tau}$ (on the reverse chain) such that $\rho = \hat{\tau}^*$. A distribution $\tau$ is *self-contrasting* if $\tau = \tau^* = \hat{\tau}^*$.

Singletons are the simplest examples of contrasted distributions: in fact, they are self contrasting. Using $\hat{i}$ to denote the singleton distribution on the reverse chain, we have $i^* = \hat{i}$ and $\hat{i}^* = i$. Indeed, the rule "take one step from $j$ and then continue walking until you reach $i$" is an optimal rule if and only if $j \neq i$.

It is far more common to have $\tau \neq \tau^*$. Moreover, for a non-reversible chain, we usually have $\tau^* \neq \hat{\tau}^*$. We are particularly interested in the case $\tau = \pi$. We have $\hat{\mu} = \pi^*$ and $\mu = \hat{\pi}^*$. We will continually remind the reader that $\hat{\mu} = \pi^*$ for compatibility with Theorem 1.2.

## 2.2 Duality of exit frequency matrices

We prove our main result. The following theorem reveals that the duality present in Theorem 1.2 is part of a more general framework. We denote column and row vectors by $\mathbf{v}$ and $\mathbf{v}^\top$, respectively. Also, we use the notation $(a_k)_{k \in S}$ to denote a column vector whose $k$th entry is $a_k$. For example, $(H(k, \pi))_{k \in S}$ denotes the column vector of access times from singletons to $\pi$.

**Theorem 2.4** *Consider a distribution $\tau$ on the forward chain. Let $X_\tau$ be the exit frequency matrix for an optimal forward $\tau$-family.*

(a) The matrix $RX_\tau^\top R^{-1}$ is the reverse exit frequency matrix for a (not necessarily optimal) reverse $\tau^*$-family.

(b) The reverse optimal exit frequency matrix for $\tau^*$ is

$$\hat{X}_{\tau^*} = R\left(X_\tau^\top - \mathbf{b}\mathbf{1}^\top\right)R^{-1} \quad \text{where} \quad \mathbf{b} = \left(\min_{i \in S} x_k(i, \tau)\right)_{k \in S}.$$

Theorem 2.4 (in analogy with (1)) states that $RX_\tau^\top R^{-1}$ is nearly equal to the reverse optimal exit frequency matrix for $\tau^*$. The rules in (a) may not be optimal. Indeed, every row of $X_\tau$ contains at least one zero, but there may be columns of $X_\tau$ which do not contain any zeros. Equivalently, there may be rows of $RX_\tau^\top R^{-1}$ without halting states. As per statement (b), we find $\hat{X}_{\tau^*}$ by subtracting the smallest entry in each row of $X_\tau^\top$ from all the entries in that row before scaling, guaranteeing that there is a zero in every row (and every column) of $\hat{X}_{\tau^*}$.

**Proof.** The $ij$th entry of $(I - M)X_\tau$ is

$$x_j(i, \tau) - \sum_{k \in S} p_{ik} x_j(k, \tau)$$

$$= \pi_j \left( H(i, \tau) - \sum_{k \in S} p_{ik} H(k, \tau) - H(i, j) + \sum_{k \in S} p_{ik} H(k, j) \right)$$

$$= i_j - \frac{\pi_j \tau_i^*}{\pi_i},$$

where $i_j$ denotes the $j$th component of the singleton distribution $i$. Indeed, we have $1/\pi_i = \text{Ret}(i) = 1 + \sum_{k \in S} p_{ik} H(k, i)$, and therefore $(I - M)X_\tau = I - R\tau^*\mathbf{1}^\top R^{-1}$. This proves statement (a).

We now prove statement (b). Rewriting the reverse conservation equation for $\tau^*$ gives $\tau^*\mathbf{1}^\top = I - (I - \hat{M}^\top)\hat{X}_{\tau^*}^\top$. Substituting this value into the previous equation and using $\hat{M}^\top = R^{-1}MR$ yields

$$(I - M)X_\tau = (I - M)R\hat{X}_{\tau^*}R^{-1}.$$

Hence $R\hat{X}_{\tau^*}^\top R^{-1} = X_\tau - L$ where $(I - M)L = 0$. The rank of $I - M$ is $n - 1$ and $\mathbf{1}$ is a right eigenvector for 0, so $L = \mathbf{b}\mathbf{1}^\top$ for some constant vector $\mathbf{b}$. Since every row of $\hat{X}_{\tau^*}$ must contain a zero element, we must have $b_k = \min_{i \in S} x_k(i, \tau)$ for $1 \leq k \leq n$. □

This theorem reveals that the forward contrast distribution $\tau^*$ has a natural interpretation on the reverse chain. Henceforth, the symbol $\tau^*$ will refer to a (forward contrasted) distribution on the reverse chain. Similarly $\hat{\tau}^*$ (the reverse contrast distribution of $\hat{\tau}$) will refer to a

11

distribution on the forward chain. In other words, a "starred" distribution is associated with the reverse chain of the original distribution.

A natural question arises: what happens if we apply these contrast maps in succession? Indeed, if we apply the reverse contrast map to $\tau^*$, we get a distribution associated with the forward chain. In general, $\tau \neq \hat{c}^*(\tau^*)$. However, the next corollary will show that these distributions are closely related. The distribution $\hat{c}^*(\tau^*)$ merits a definition.

**Definition 2.5** Let $\tau$ be a distribution on the forward chain. The *forward $\tau$-core* distribution is given by $\tau^{**} = \hat{c}^*(c^*(\tau)) = \hat{c}^*(\tau^*)$. Let $\hat{\tau}$ be a distribution on the reverse chain. The *reverse $\hat{\tau}$-core* distribution is given by $\hat{\tau}^{**} = c^*(\hat{c}^*(\hat{\tau})) = c^*(\hat{\tau}^*)$.

The following corollary shows that the distribution $\tau^{**}$ is fully dual to $\tau^*$.

**Corollary 2.6** *For any distribution $\tau$, we have*

$$X_{\tau^{**}} = R\hat{X}_{\tau^*}^\top R^{-1}.$$

*Furthermore, $c^*(\tau^{**}) = \tau^*$ and $\hat{c}^* \circ c^*(\tau^{**}) = \tau^{**}$.*

**Proof.** Applying Theorem 2.4 to the reverse chain,

$$X_{\tau^{**}} = R\left(\hat{X}_{\tau^*}^\top - \mathbf{a}\mathbf{1}^\top\right)R^{-1} \quad \text{where} \quad \mathbf{a} = \left(\min_{i \in S} \hat{x}_k(i, \tau^*)\right)_{k \in S}.$$

Each column of $\hat{X}_{\tau^*}$ contains at least one zero entry, and therefore $\left(\min_{i \in S} \hat{x}_k(i, \tau^*)\right)_{k \in S} = \mathbf{0}$. In addition, each row of $\hat{X}_{\tau^*}$ contains at least one zero. (Indeed, $\hat{X}_{\tau^*} \leq X_\tau$ and $X_\tau$ is an optimal exit frequency matrix.) These two facts give $c^*(\tau^{**}) = \tau^*$. $\square$

The relationship described in corollary 2.6 can be summarized via a formula using scaled exit frequencies:

$$\hat{y}_i(j, \tau^*) = y_j(i, \tau^{**}) \text{ for all } i, j \in S. \tag{22}$$

It is also important to note that both $\tau^*$ and $\tau^{**}$ are "balanced" in the sense that every state is a halting state for at least one other state. Furthermore, equation (22) implies that $i$ is a reverse $(j, \tau^*)$-halting state if and only if $j$ is a forward $(i, \tau^{**})$-halting state. We recognize the perfect duality of $\tau^*$ and $\tau^{**}$ with the following terminology.

**Definition 2.7** Let $\tau$ be a distribution on the forward chain. The ordered pair of distributions $(\tau^*, \tau^{**})$ is called a *(forward) $\tau$-contrasting pair*. More generally, the ordered pair of

distributions $(\hat{\rho}, \sigma)$ is a (forward) contrasting pair if there exists a distribution $\tau$ such that $\hat{\rho} = \tau^*$ and $\sigma = \tau^{**}$. The analogous definitions for *reverse contrasting pairs* $(\hat{\tau}^*, \hat{\tau}^{**})$ and $(\rho, \hat{\sigma})$ are self evident.

The contrasting pair $(\pi^*, \pi^{**})$ will play a central role in our new proofs of Theorems 1.1 and 1.2 in Section 3.

**Corollary 2.8** *The pair of distributions $(\hat{\rho}, \sigma)$ are a contrasting pair if and only if $X_\sigma = R\hat{X}_{\hat{\rho}}^\top R^{-1}$.*  $\square$

Our last result in this section concerns distributions that are "on the way" to $\tau$. Let $W_\tau = \{\sigma : \forall i \in S, H(i, \sigma) + H(\sigma, \tau) = H(i, \tau)\}$. This set is nonempty: $\tau \in W_\tau$ since $H(\tau, \tau) = 0$. We interpret $W_\tau$ as the set of distributions "on the way" to $\tau$ in the sense that the rule "walk from $i$ to $\sigma$ and then from $\sigma$ to $\tau$" is optimal for each singleton $i$. In other words, you may choose to obtain a sample from $\sigma$ on your way to obtaining a sample from $\tau$ optimally.

**Corollary 2.9** *The $\tau$-core $\tau^{**}$ is the unique reverse contrasted distribution in $W_\tau$. Moreover, $H(\tau^{**}, \tau) = \max_{\sigma \in W_\tau} H(\sigma, \tau)$.*

**Proof.** First, we show that $\tau^{**} \in W_\tau$. The rule "follow an optimal rule from $i$ to $\tau^{**}$ and then follow an optimal rule from $\tau^{**}$ to $\tau$" has a halting state. Indeed, we have $x_j(i, \tau^{**}) \leq x_j(i, \tau)$ for all $i, j \in S$. If $k$ is an $(i, \tau)$-halting state then it must also be an $(i, \tau^{**})$-halting state, so $x_k(i, \tau^{**}) + x_k(i, \tau) = 0$. This composite rule has a halting state, so it is optimal. Therefore we have $H(i, \tau) = H(i, \tau^{**}) + H(\tau^{**}, \tau)$.

Next, consider any $\sigma \in W_\tau$. Since $H(i, \tau) = H(i, \sigma) + H(\sigma, \tau)$ for all $i$, we have $x_k(i, \sigma) + x_k(\sigma, \tau) = x_k(i, \tau)$. The exit frequency $x_k(\sigma, \tau)$ is independent of $i$ and all exit frequencies must be nonnegative. Therefore, $x_k(i, \sigma) \geq x_k(i, \tau) - \min_{j \in S} x_k(j, \tau) = x_k(i, \tau^{**})$ for all $i, k$. So $H(i, \sigma) \geq H(i, \tau^{**})$ for all $i$, which in turn guarantees that $H(\sigma, \tau) \leq H(\tau^{**}, \tau)$.

Finally, $\tau^{**}$ is the unique reverse contrasted distribution among all $\sigma$ such that $H(i, \tau) = H(i, \sigma) + H(\sigma, \tau)$ for all $i$. Indeed, we have $X_\sigma = X_\tau - \mathbf{1}\mathbf{c}^\top$ for the vector $\mathbf{c}$ given by $c_k = x_k(\sigma, \tau)$. In order for $\sigma$ to be reverse contrasted, we must have a zero in every column of $X_\sigma$. We obtain a zero in the $k$th column if and only if $c_k = \min_{j \in S} x_k(j, \tau)$.  $\square$

In closing, we reflect on the characteristics contrasting pairs, and underscore our naming convention for these distributions. The $\tau$-contrast $\tau^*$ has a natural association with the reverse chain. Yet its also captures the *differences* between forward rules from neighboring states to $\tau$: $\tau_i^*$ is nonzero when $i$ and all of its neighbors do not share a single common halting state. On the other hand, the $\tau$-core $\tau^{**}$ has a natural association with the forward chain. In addition, the distribution $\tau^{**}$ captures the *commonality* among all forward rules from singletons to $\tau$: $\tau^{**}$ is the "furthest" distribution from $\tau$ from which you can sample optimally on your way to sampling from $\tau$ starting from any singleton.

## 2.3 Examples of contrasting pairs

We illustrate contrast and core distributions using random walks on graphs. An undirected graph corresponds to a time reversible Markov chain, in which case $\tau^* = \hat{\tau}^*$ and $\tau^{**} = \hat{\tau}^{**}$. Since every singleton distribution is self contrasting, we focus on nonsingleton distributions, with special attention to the forget distribution $\mu = \pi^*$ and the $\pi$-core $\pi^{**}$.

We consider two undirected graph structures: complete graphs and trees. These graphs are simple enough that we can completely characterize their contrasting pairs. We also consider one directed graph: the winning streak (introduced in [12]).

**Complete Graph.** Let $\tau$ be any distribution on $K_n$ with nodes ordered so that $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_n$. Equation (16) and $H(i,j) = n - 1$ for $i \neq j$ guarantee that node 2 is a $(1, \tau)$-halting state and node 1 is a $(i, \tau)$-halting state for $i > 1$. By Theorem 2.4, $RX_{\tau^*}^\top R^{-1} = X_\tau - \mathbf{1b}^\top$ where $b_k = \min_{i \in S} x_k(i, \tau)$. By simply transposing $X_\tau$ it is clear that every node $i > 1$ is a $(1, \tau^*)$-halting state, so $\tau_2^* = \tau_3^* = \cdots = \tau_n^*$. Moreover, $x_1(1, \tau^*) = x_1(1, \tau) = \frac{n-1}{n}(1 - \tau_1 + \tau_2)$. By equation (20), $\tau_1^* = \frac{1}{n} + \frac{n-1}{n}(\tau_1 - \tau_2)$ and $\tau_k^* = \frac{1}{n}(1 - \tau_1 + \tau_2)$ for $2 \leq k \leq n$.

Let $\tau$ be a non-singleton distribution on the complete graph $K_n$ ordered so that $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_n$. It follows from the above that $\tau$ is a contrasted distribution if any only if $\tau_2 = \cdots = \tau_n$. Note that every contrasted distribution on $K_n$ is self-contrasting: $\tau^* = \tau^{**}$, and therefore $H(i, \tau^*) + H(\tau^*, \tau) = H(i, \tau)$ for all $i$ by corollary 2.6.

Turning our attention to the contrasting pair for $\pi$, the symmetry of the graph gives $H(i, \pi) = H(j, \pi)$ all $i, j \in S$, and hence $\mu = \pi^* = \pi^{**} = \pi$.

**Trees.** Stopping rules on trees were studied extensively in [3], which introduced the notion of a *focus* of a distribution. Given a target distribution $\tau$, a $\tau$-*focus* is a node $i$ such that

$H(i, \tau) < 1 + \sum_{j \in V(T)} p_{ij} H(j, \tau)$. In other words, the foci of $\tau$ correspond to the support of $\tau^*$: $S_{\tau^*} = \{i \mid \tau_i^* > 0\}$. Theorem 5 in [3] states that every distribution either has one focus or two adjacent foci. This guarantees that every contrasted distribution on a tree must either be a singleton or concentrated on two adjacent nodes. Like the complete graph, each of these is self-contrasting: $\tau^* = \tau^{**}$.

In particular, we consider $\tau = \pi$ for a path on $n$ vertices and calculate $\mu = \pi^* = \pi^{**}$. In this case, the best choice to "forget" where we started is to walk to the center of the path. For a path on $2k + 1$ nodes, $\mu$ will be concentrated on node $k + 1$, the unique the center of the path. For a path on $2k$ nodes, $\mu$ will be evenly divided between the two central nodes $k$, $k + 1$ of the path. The following general formula for $\mu$ on a tree is given in proposition 7 of [3]. If $\pi$ has a single focus, then $\mu$ is a singleton. If $\pi$ has two adjacent foci $a, b \in V(T)$ then $\mu_a = (H(b', b) - H(a', b))/2|E|$ and $\mu_b = (H(a', a) - H(b', a))/2|E|$, where $a'$ and $b'$ are pessimal nodes as defined in equation (18).

**Winning Streak.** The winning streak chain nicely illustrates how distinct our mixing walks can be for the forward and reverse chain. The winning streak is a finite digraph whose random walk behaves much like a gambler who "lets it ride" on each subsequent bet. We have chosen to loops on the extreme nodes on this graph because it leads to a nice stationary distribution whose components are all powers of $1/2$.

The winning streak on $n$ nodes $\{0, 1, \ldots, n - 1\}$ has transition probabilities given by

$$p_{ij} = \begin{cases} 1/2 & j = i + 1 \\ 1/2 & j = 0 \text{ and } 0 \leq i \leq n - 1 \\ 1/2 & i = j = n - 1 \\ 0 & \text{otherwise.} \end{cases}$$

The stationary distribution is $\pi = (1/2, 1/4, \ldots, 2^{-n+1}, 2^{-n+1})$. Both state $0$ and $n - 1$ achieve $T_{\mathrm{mix}}$. Indeed, for either one, the trivial rule "take $n - 1$ steps" is an optimal rule achieving $\pi$.

The transition probabilities for the reverse winning streak are

$$\hat{p}_{ij} = \begin{cases} 1 & j = i - 1 \text{ and } 1 \leq i \leq n - 2 \\ 2^{-j-1} & i = 0 \text{ and } 0 \leq j \leq n - 2 \\ 2^{-n+1} & i = 0 \text{ and } j = n - 1 \\ 1/2 & i = n - 1 \text{ and } j = n - 1 \text{ or } j = n - 2 \\ 0 & \text{otherwise.} \end{cases}$$

For the reverse winning streak, state $n - 1$ achieves $\hat{T}_{\mathrm{mix}}$: the rule "take $n - 1$ steps" is an optimal mixing rule from this state. However, node $0$ is an advantageous starting point: the (optimal) rule "take one step" takes us immediately from node $0$ to $\pi$.

We describe the forget distribution and $\pi$-core of both the forward and reverse winning streak chains, omitting the calculation details. The reverse forget distribution was calculated in [12]:

$$\mu = \hat{\pi}^* = \begin{cases} 1 - 2^{-n+1} & i = 0 \\ 2^{-n+1} & i = n-1 \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $H(i, \mu) = 2 - 2^{-n+2}$ for all $i$ (walk until either state 0 is hit or $n-1$ steps have been made) and therefore $\hat{\pi}^{**} = c^*(\mu) = \mu^* = \hat{\pi}$. In other words, $\hat{\pi}$ is forward contrasted.

The contrasting pair $\hat{\mu} = \pi^*$ and $\pi^{**}$ are not as well behaved. A more complicated analysis of the mixing walks on the forward chain shows that if $i$ is among the first $\log_2 n$ nodes, then node $n-1$ is an $(i, \pi)$-halting state. Each remaining node $i$ has node $i-1$ as an $(i, \pi)$-halting state. More precisely, let $i_0$ be the unique node satisfying $2^{i_0} + i_0 \leq n < 2^{i_0+1} + i_0 + 1$. Some routine (but lengthy) calculations of access times gives

$$\hat{\mu} = \pi^* = \begin{cases} 0 & 0 \leq i < i_0 \\ \frac{\pi_{i_0}}{2}(2^{i_0+2} - n + i_0) & i = i_0 \\ \frac{\pi_{i_0}}{2}(2 + n - i) & i_0 < i \leq n - 2 \\ \pi_{n-1} & i = n - 1. \end{cases}$$

Finally, we may calculate $\pi^{**} = \hat{\mu}^*$ using Theorem 2.4:

$$\pi^{**} = \begin{cases} \sum_{k=0}^{i_0} \pi_k & i = 0 \\ 0 & 1 \leq i \leq i_0 \\ \pi_i & i_0 < i \leq n - 1. \end{cases}$$

In summary, on the winning streak, $\pi^*, \pi^{**}, \hat{\pi}^*$, and $\hat{\pi}^{**}$ are quite different from one another.

## 2.4   Duality of contrasting pairs

In this section, we further explore the duality of contrasting pairs. We begin by generalizing the notion of state-to-state commute time $\kappa(i, j) = H(i, j) + H(j, i)$ to distributions by $\kappa(\sigma, \tau) = H(\sigma, \tau) + H(\tau, \sigma)$. We have the following extension of equation (11).

**Proposition 2.10** *For a contrasting pair* $(\tau^*, \tau^{**})$ *and any state* $i$, $\hat{\kappa}(i, \tau^*) = \kappa(i, \tau^{**})$.

**Proof.**    The equality $y_i(i, \tau^{**}) = \hat{y}_i(i, \tau^*)$ of equation (22) is equivalent to $H(i, \tau^{**}) + H(\tau^{**}, i) = \hat{H}(i, \tau^*) + \hat{H}(\tau^*, i)$.                                   $\square$

Furthermore, the average access times to these distributions are equal.

**Proposition 2.11** *The contrasting pair* $(\tau^*, \tau^{**})$ *satisfy*

$$\sum_{k \in S} \pi_k \hat{H}(k, \tau^*) = \sum_{k \in S} \pi_k H(k, \tau^{**}) \leq \sum_{k \in S} \pi_k H(k, \tau).$$

**Proof.** Taking the trace of both $\hat{X}_{\tau^*}$ and $R X_{\tau^{**}}^\top R^{-1}$,

$$\sum_{k \in S} \pi_k (\hat{H}(k, \tau^*) + \hat{H}(\tau^*, k)) = \sum_{k \in S} \pi_k (H(k, \tau^{**}) + H(\tau^{**}, k))$$

therefore $\sum_{k \in S} \pi_k \hat{H}(k, \tau^*) = \sum_{k \in S} \pi_k H(k, \tau^{**})$ by the random target identity (13). Since $H(k, \tau^{**}) \leq H(k, \tau)$ for all $k$, we have $\sum_{k \in S} \pi_k H(k, \tau^{**}) \leq \sum_{k \in S} \pi_k H(k, \tau)$. $\qquad\square$

A third duality result between a contrasting pair concerns the regeneration time, which was studied in [4]. The *regeneration time* of a distribution $\rho$ is the expected time between independent samples:

$$T_{\text{regen}} = \sum_{i \in S} \rho_i H(i, \rho).$$

**Theorem 2.12** *For a distribution* $\tau$,

$$\hat{T}_{\text{regen}}(\tau^*) = T_{\text{regen}}(\tau^{**}) \leq T_{\text{regen}}(\tau).$$

**Proof.** Given $\tau$, we prove that $\hat{T}_{\text{regen}}(\tau^*) \leq T_{\text{regen}}(\tau)$. First, we observe that Theorem 2.4(b) ensures that $y_k(i, \tau) \geq \hat{y}_i(k, \tau^*)$ for every $i, k \in S$. Pick any $i \in S$ and then choose a $(i, \tau)$-halting state $j$. This means that $y_j(i, \tau) = 0$ and therefore $\hat{y}_i(i, t^*) = 0$ as well. For this choice of $i, j$, we applying equation (10) with $\sigma = \tau$ and $\hat{\tau} = \tau^*$. This gives

$$
\begin{aligned}
-\hat{y}_i(j, \tau^*) + \sum_{k \in S} \tau_k \hat{y}_k(j, \tau^*) &= -y_j(i, \tau) + \sum_{k \in S} \tau_k^* y_k(i, \tau) \\
\sum_{k \in S} \tau_k y_j(k, \tau) &\geq \sum_{k \in S} \tau_k^* \hat{y}_i(k, \tau^*) \\
\sum_{k \in S} \tau_k (H(k, \tau) + H(\tau, j) - H(k, j)) &\geq \sum_{k \in S} \tau_k^* (\hat{H}(k, \tau^*) + \hat{H}(\tau^*, i) - \hat{H}(k, i)) \\
\sum_{k \in S} \tau_k H(k, \tau) &\geq \sum_{k \in S} \tau_k^* \hat{H}(k, \tau^*).
\end{aligned}
$$

The third line follows from equation (7). The final line states that $T_{\text{regen}}(\tau) \geq \hat{T}_{\text{regen}}(\tau^*)$.

The equality $T_{\text{regen}}(\tau^{**}) = \hat{T}_{\text{regen}}(\tau^*)$ follows from duality. Replacing $\tau$ with $\tau^{**}$ in the above argument gives $T_{\text{regen}}(\tau^{**}) \geq \hat{T}_{\text{regen}}(\tau^*)$. We then make the parallel argument, swapping

the roles of the forward and reverse chains to find that $T_{\text{regen}}(\tau^{**}) \leq \hat{T}_{\text{regen}}(\tau^*)$, so indeed we have equality. $\qquad\square$

The next two characterizations of halting states follow immediately from Proposition 2.2 and Theorem 2.4.

**Proposition 2.13** *A state $z$ is a reverse $(\pi, \tau^*)$-halting state if and only if $z$ achieves $\min_{k \in S} H(k, \tau)$; equivalently, if it achieves $\min_{k \in S} H(k, \tau^{**})$. Moreover,*

$$\min_{k \in S} H(k, \tau) = \sum_{i \in S} \pi_i H(i, \tau) - \hat{H}(\pi, \tau^*),$$

*and*

$$\min_{k \in S} H(k, \tau^{**}) = \sum_{i \in S} \pi_i \hat{H}(i, \tau^*) - \hat{H}(\pi, \tau^*).$$

$\qquad\square$

**Proposition 2.14** *A state $z$ is a reverse $(\tau^*, \pi)$-halting state if and only if $z$ achieves $\max_{k \in S} H(k, \tau)$; equivalently, if it achieves $\max_{k \in S} H(k, \tau^{**})$. Moreover,*

$$\max_{k \in S} H(k, \tau) = \sum_{i \in S} \pi_i H(i, \tau) + \hat{H}(\tau^*, \pi),$$

*and*

$$\max_{k \in S} H(k, \tau^{**}) = \sum_{i \in S} \pi_i \hat{H}(i, \tau^*) + \hat{H}(\tau^*, \pi).$$

$\qquad\square$

Finally, we extend the cycle reversing identity (12) for contrasting pairs.

**Proposition 2.15** *The pair of distributions $(\hat{\rho}, \sigma)$ is a contrasting pair if and only if*

$$\hat{H}(i, \hat{\rho}) + \hat{H}(\hat{\rho}, j) + \hat{H}(j, i) = H(j, \sigma) + H(\sigma, i) + H(i, j)$$

*for all $i, j \in S$.*

**Proof.** Suppose that the above equalities hold for the pair $(\hat{\rho}, \sigma)$. By By equation (11), we may subtract $\hat{\kappa}(i, j)$ from the left hand side and $\kappa(i, j)$ from the right hand side to obtain $\hat{y}_j(i, \hat{\rho}) = y_i(j, \sigma)$ for all $i, j \in S$. By Corollary 2.8, this hold if and only if $(\hat{\rho}, \sigma)$ are a contrasting pair. $\qquad\square$

It follows from this proposition that $\tau$ is self contrasting if and only if

$$\hat{H}(i,\tau) + \hat{H}(\tau,j) + \hat{H}(j,i) = H(j,\tau) + H(\tau,i) + H(i,j) \text{ for all } i,j \in S. \qquad (23)$$

# 3   Mixing measures

We apply the framework of exit frequency matrices and contrasting pairs to give new proofs of Theorems 1.1 and 1.2. For the equality $\hat{T}_{\text{forget}} = T_{\text{reset}}$ of Theorem 1.2, new proof proof offers insight into the relationship between $\pi$ and the reverse forget distribution $\hat{\mu} = \pi^*$.

**Proof of Theorem 1.2.**    If $\sigma$ achieves $\min_\tau \max_{i \in S} \hat{H}(i,\tau))$ then $\sigma$ must be contrasted: we have $\hat{x}_k(i,\hat{\sigma}^{**}) \le \hat{x}_k(i,\sigma)$ by Theorem 2.4, so $\hat{H}(i,\hat{\sigma}^{**}) \le \hat{H}(i,\sigma)$, and so $\sigma = \hat{\sigma}^{**}$. Thus we may assume that our minimum is achieved by $\tau^*$ where $(\tau^*,\tau^{**})$ are a contrasting pair. Applying corollary 2.14 for the reverse chain (and switching the roles of $\tau^*$ and $\tau^{**}$),

$$\max_{i \in S} \hat{H}(i,\tau^*) = \sum_{j \in S} \pi_j (H(j,\tau^{**}) + H(\tau^{**},\pi)) \ge \sum_{j \in S} \pi_j H(j,\pi) = T_{\text{reset}}.$$

Using corollary 2.6 with $\tau = \pi$, we find that $H(j,\pi^{**}) + H(\pi^{**},\pi) = H(j,\pi)$ for all $j$ and therefore $\pi^* = \hat{\mu}$ achieves this lower bound. The uniqueness of $\hat{\mu}$ also follows from corollary 2.6. $\qquad\square$

We now use Theorem 1.2 to prove Theorem 1.1, showing that $T_{\text{mix}} = \hat{T}_{\text{mix}}$.

**Proof of Theorem 1.1.**   By corollary 2.14 with $\tau = \pi$ and $\tau^* = \hat{\mu}$, $T_{\text{mix}} = \max_{k \in S} H(k,\pi) = T_{\text{reset}} + \hat{H}(\hat{\mu},\pi) = \hat{T}_{\text{forget}} + \hat{H}(\hat{\mu},\pi) = \max_{k \in S}(\hat{H}(k,\hat{\mu}) + \hat{H}(\hat{\mu},\pi)) \ge \max_{k \in S} H(k,\pi) = \hat{T}_{\text{mix}}$. Starting with the mixing time for the reverse chain, we similarly find that $\hat{T}_{\text{mix}} \ge T_{\text{mix}}$ so equality must hold everywhere. $\qquad\square$

**Corollary 3.1**  *If $H(z,\pi) = \max_i H(i,\pi)$ then $\hat{H}(z,\hat{\mu}) + \hat{H}(\hat{\mu},\pi) = \hat{H}(z,\pi)$.* $\qquad\square$

**Corollary 3.2**  *$\hat{H}(z,\hat{\mu}) = \max_i \hat{H}(i,\hat{\mu})$ if and only if there is a state $i \in S$ such that $z$ is an $(i,\pi)$-halting state on the forward chain.*

**Proof.**   Summing across the $i$th row of $\hat{X}_{\hat{\mu}}$ gives

$$\hat{H}(i,\hat{\mu}) = \sum_{j \in S} \pi_j \hat{y}_j(i,\hat{\mu}) = \sum_{j \in S} \pi_j y_i(j,\pi) - \min_{k \in S} y_i(k,\pi) = \sum_{j \in S} \pi_j H(j,\pi) - \min_{k \in S} y_i(k,\pi)$$

19

which is maximized whenever $\min_{k\in S} y_i(k,\pi) = 0$. $\qquad\square$

This is an improvement over Corollary 7.1 of [12] which proved that if $H(z,\pi) = \max_i H(i,\pi)$ then $\hat{H}(z,\hat{\mu}) = \max_i \hat{H}(i,\hat{\mu})$. Finally, our next corollary is dual to Theorem 6.1 of [12], which states that $T_{\text{forget}} = \sum_{j\in S} \pi_j(H(j',j) - H(\pi,j))$ where $\hat{H}(j',j) = \max_{k\in S} \hat{H}(k,j)$.

**Corollary 3.3** $\hat{T}_{\text{forget}} = \sum_{j\in S} \hat{\mu}_j(\hat{H}(j',j) - \hat{H}(\hat{\mu},j))$ *where* $\hat{H}(j',j) = \max_{k\in S} \hat{H}(k,j)$.

**Proof.** Consider any $j \in S$. Since $j'$ is a $(j,\pi)$-halting state, Theorem 2.4(b) shows that $j$ is a reverse $(j',\hat{\mu})$-halting state. Proposition 3.2 guarantees that $\hat{H}(j',j) - \hat{H}(\hat{\mu},j) = \hat{H}(j',\hat{\mu}) = \hat{T}_{\text{forget}}$ for all $j$. $\qquad\square$

# 4 Structure of exit frequency matrices

## 4.1 Inverting optimal exit frequency matices

Not surprisingly, the exit frequency matrix $X_\tau$ is closely related to $I - M$. The rank of $I - M$ is $n-1$ ($\pi$ is a left eigenvector for eigenvalue 0). Depending on our target distribution, a slight alteration of $I - M$ gives the inverse of $X_\tau$.

**Theorem 4.1** *Consider a singleton target distribution* $\tau = k$ *with optimal exit frequency matrix* $X_k$. *Let* $X'_k$ *be the* $(n-1)\times(n-1)$ *matrix obtained by deleting the $k$th row and the $k$th column of* $X_k$. *Let* $M'$ *be obtained similarly from the transition matrix* $M$. *Then* $X'_k$ *is the inverse of* $(I - M')$.

Note that when our target is the singleton $\tau = k$ then the $k$th row and the $k$th column are both zero, hence $n-1$ is the largest possible rank for $X_k$.

**Proof.** Fix a target singleton distribution $k$ and let $X_k = (x_j(i,k))_{i,j\in S}$ be the exit frequency matrix of an optimal family of rules from the singletons to $k$. The $k$th row and $k$th column are both 0, and therefore the rank of $X_k$ is at most $n-1$. Let $X'_k$ and $M'$ be the matrices defined by deleting the $k$th row and column from $X_k$ and $M$, respectively. Consider the $ij$th entry of $X'_k(I - M')$ which is

$$x_j(i,k) - \sum_{r\in S\setminus\{k\}} p_{rj}x_r(i,k) = i_j$$

20

by the conservation equation (4). Hence $X'_k(I - M') = I$. Similarly, we see that the $ij$th entry of $(I - M')X'_k$ satisfies

$$x_j(i, k) - \sum_{r \in S \backslash \{k\}} p_{ir} x_j(r, k) = i_j$$

by considering the optimal rule "take one step from $i$ and follow an optimal rule to $k$." Thus, $X'_k$ is the inverse of $I - M'$ (which must have rank $n - 1$). □

We note that this theorem is equivalent to a special case of the so-called *fundamental matrix* described by Doyle and Snell on page 27 of [6]. Treating our target state as an absorbing state, the resulting transition matrix $P$ is of the form

$$P = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{pmatrix}$$

where $Q$ is an $(n-1) \times (n-1)$ square matrix. The fundamental matrix for this absorbing chain is $N = (I - Q)^{-1} = \sum_{t=0}^{\infty} Q^t$. Doyle and Snell observe that $N_{ij}$ is the expected number of times that the chain will be in state $i$ before absorption when it is started in state $j$. This is equivalent to our exit formulation since every time we visit $i \neq k$, we must also exit.

When $\tau$ is not a singleton, the optimal exit frequency matrix $X_\tau$ has full rank. We obtain $X_\tau^{-1}$ by adding the appropriate rank 1 matrix to $I - M$.

**Theorem 4.2** *If $\tau$ be a target distribution that is not concentrated on a single state, then*

$$X_\tau^{-1} = I - M + \frac{1}{T_{\text{regen}}(\tau)} R \tau^* \tau^\top. \tag{24}$$

**Proof.** The $ij$th entry of the matrix product $\left( I - M + \frac{1}{T_{\text{regen}}(\tau)} R \tau^* \tau^\top \right) X_\tau$ is

$$x_j(i, \tau) - \sum_{k \in S} p_{ik} x_j(k, \tau) + \frac{1}{T_{\text{regen}}(\tau)} \frac{\tau_i^*}{\pi_i} \sum_{k \in S} \tau_k x_j(k, \tau).$$

$$= \pi_j \left( H(i, \tau) - \sum_{k \in S} p_{ik} H(k, \tau) - H(i, j) + \sum_{k \in S} p_{ik} H(k, j) + \frac{\tau_i^*}{\pi_i} \right)$$

by equation (7). Considering the case $i \neq j$, we have $H(i, j) - \sum_{k \in S} p_{ik} H(k, j) = 1$ and the $ij$th entry of equation is

$$\pi_j \left( H(i, \tau) - \sum_{k \in S} p_{ik} H(k, \tau) - 1 + \frac{\tau_i^*}{\pi_i} \right) = \pi_j \left( -\frac{\tau_i^*}{\pi_i} + \frac{\tau_i^*}{\pi_i} \right) = 0$$

21

by equation (20). The $i$th diagonal entry is

$$\pi_i \left( H(i,\tau) - \sum_{k \in S} p_{ik} H(k,\tau) + \sum_{k \in S} p_{ik} H(k,i) + \frac{\tau_i^*}{\pi_i} \right) = \pi_i \left( \frac{1 - \tau_i^*}{\pi_i} + \frac{\tau_i^*}{\pi_i} \right) = 1$$

since $1 + \sum_{k \in S} p_{ik} H(k,i) = \mathrm{Ret}(i) = 1/\pi_i$ is the return time to $i$. $\qquad\square$

As defined in equation (15), the naive access time $N(\sigma, \tau)$ is the expected length of a naive rule from $\sigma$ to $\tau$ (on the forward chain) .

**Corollary 4.3** *For any distribution $\tau$ and any state $i \in S$,*

$$T_{\mathrm{regen}}(\tau) - N(\tau, \tau^*) = H(i,\tau) - N(i, \tau^*).$$

Simply put, this corollary states that the quantity $H(i,\tau) - N(i,\tau^*)$ is independent of the initial state $i$. This corollary is surprising: it concerns non-optimal $\tau^*$-families on the *forward* chain (rather than the reverse chain).

**Proof.**   If $\tau$ is a singleton, then the result is trivial: a singleton is self-contrasting and the naive rule to a singleton is optimal. Considering a non-singleton $\tau$, the proof is a straight forward expansion of $X_\tau X_\tau^{-1} = I$. Indeed, the matrix conservation equation (9) gives

$$
\begin{aligned}
X_\tau \left( I - M + \frac{1}{T_{\mathrm{regen}}(\tau)} R\tau^* \tau^\top \right) &= I \\
I - \mathbf{1}\tau^\top + \frac{1}{T_{\mathrm{regen}}(\tau)} X_\tau R \tau^* \tau^\top &= I \\
Y_\tau \tau^* \tau^\top &= T_{\mathrm{regen}}(\tau) \, \mathbf{1}\tau^\top.
\end{aligned}
$$

Choose $j \in S$ so that $\tau_j \neq 0$. For each $i \in S$, $ij$th entry gives the equality $\sum_{k \in S} \tau_k^* y_k(i,\tau) = \sum_{k \in S} \tau_k^* \big( H(i,\tau) + H(\tau,k) - H(i,k) \big) = T_{\mathrm{regen}}(\tau)$ where we use equation (7) to expand the exit frequencies. $\qquad\square$

In the important case where $\tau = \pi$, we rewrite this equation using our mixing measure notation:

$$X_\pi^{-1} = I - M + \frac{1}{T_{\mathrm{reset}}} R\hat{\mu}\pi^\top.$$

and

$$T_{\mathrm{reset}} - N(\pi, \hat{\mu}) = H(i,\pi) - N(i,\hat{\mu}).$$

The latter equation gives a particularly interesting result for a random walk on a tree.

22

**Corollary 4.4** *For any nontrivial tree, $T_{\text{mix}} < 2\,T_{\text{reset}}$.*

This gives a tighter bound for trees than the known bound $T_{\text{mix}} \leq 4\,T_{\text{reset}}$ for a time reversible Markov chain (see [2]).

**Proof.**    A random walk on a tree is time reversible, so $T_{\text{reset}} - N(\pi, \mu) = H(i, \pi) - N(i, \mu)$. Choose $i \in S$ so that $H(i, \pi) = T_{\text{mix}}$. By Corollary 3.2 and Theorem 1.2, $H(i, \mu) = T_{\text{forget}} = T_{\text{reset}}$. As stated in Section 2.3, the forget distribution $\mu = \pi^*$ is concentrated on one node or two adjacent nodes, called foci. It is easy to see that the naive rule from $i$ to $\mu$ is optimal (the furthest focus is a halting state). Therefore $T_{\text{mix}} = 2\,T_{\text{reset}} - N(\pi, \mu) < 2\,T_{\text{reset}}$.    □

## 4.2    The spectrum of $X_\pi$

In general, the spectrum of $X_\tau$ is difficult to describe. However, we can find the spectrum of $X_\pi$ exactly.

**Theorem 4.5** *Denote the eigenvalues of $M$ by $\lambda_1 = 1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -1$. The largest eigenvalue for $X_\pi$ is $T_{\text{reset}}$ with corresponding left eigenvector $\pi$. The remaining eigenvalues are $1/(1 - \lambda_k)$ for $2 \leq k \leq n$.*

It follows immediately that

$$\frac{1}{1 - \lambda_2} \leq T_{\text{reset}}, \tag{25}$$

which was originally proven indirectly in [13]. We can confirm another spectral result by taking the trace of $X_\pi$. The trace is equal to the sum of the eigenvalues and hence

$$T_{\text{reset}} + \sum_{k=2}^{n} \frac{1}{1 - \lambda_k} = \sum_{k \in S} x_k(k, \pi) = T_{\text{reset}} + T_{\text{hit}},$$

recovering the well known identity $T_{\text{hit}} = \sum_{k=2}^{n} 1/(1 - \lambda_k)$.

**Proof of Theorem 4.5.**    Using equation (7) with the $i$th component of $\pi^\top X_\pi$ yields $\sum_{k \in S} \pi_k x_i(k, \pi) = \pi_i \sum_{k \in S} \pi_k(H(k, \pi) + H(\pi, i) - H(k, i)) = \pi_i \sum_{k \in S} \pi_k H(k, \pi) = \pi_i T_{\text{reset}}$ by the random target identity (13). Since the eigenvector $\pi > \mathbf{0}$, the Frobenius-Perron theorem (cf. [8]) proves that $T_{\text{reset}}$ is the largest eigenvalue of $X_\pi$.

Let $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$ be the respective right eigenvectors of $\lambda_2, \ldots \lambda_n$. The vector $\pi^\top$ is a left eigenvector of $M$, hence $\pi^\top \mathbf{v}_k = 0$ for $2 \leq k \leq n$. Therefore,

$$X_\pi^{-1} \mathbf{v}_k = (I - M)\mathbf{v}_k = (1 - \lambda_k)\mathbf{v}_k \text{ for } 2 \leq k \leq n,$$

and the theorem follows. □

## 4.3 Characterization of exit frequency matrices

We conclude by giving necessary and sufficient conditions for a matrix to be the exit frequency matrix for some distribution $\tau$ on a given Markov chain.

**Proposition 4.6** *Let $X$ be an $n \times n$ matrix. There exists a distribution $\tau$ such that $X$ contains the exit frequencies for a (possibly non-optimal) $\tau$-family on the Markov chain with transition matrix $M$ if and only if the following two conditions hold:*

*(a) $(I - M)X(I - M) = (I - M)$,*

*(b) $I - X(I - M) \geq 0$.*

*The matrix $X$ is the optimal exit frequency matrix for $\tau$ if and only if in addition to the above two conditions, we have*

*(c) $\min_{j \in S} X_{ij} = 0$ for $1 \leq i \leq n$.*

*The matrix $X$ is the optimal exit frequency matrix for a reverse contrasted distribution if and only if in addition to the above three conditions, we have*

*(d) $\min_{i \in S} X_{ij} = 0$ for $1 \leq j \leq n$.*

**Proof.** If $X$ is an exit frequency matrix for some family of (possibly non-optimal) rules, then (a) and (b) follow from the conservation equation (4). Considering the reverse direction, assume that $X$ satisfies both conditions. By (a), $(I-M)(X(I-M)-I) = 0$ and therefore $X(I-M) - I = -\tau \mathbf{1}^\top$ for some vector $\tau$, which verifies the conservation equation (9). Furthermore, $\sum_{k \in S} \tau_k = \tau \mathbf{1}^\top \mathbf{1} = (I - X(I - M))\mathbf{1} = 1$ and condition (b) ensures that $\tau \geq 0$, so $\tau$ is a distribution.

$X$ is an optimal exit frequency matrix if and only if each row contains a halting state, proving (c). Condition (d) follows similarly: the target distribution is reverse contrasted if and only if each column contains a zero entry. □

# References

[1] D. J. Aldous, J. Fill, Reversible Markov Chains and Random Walks on Graphs (book), to appear, `http://www.stat.berkeley.edu/~aldous/RWG/book.html`.

[2] D. Aldous, L. Lovász and P. Winkler, Mixing times for uniformly ergodic Markov chains, *Stochastic Processes and their Applications* **71** (1997), 165–185.

[3] A. Beveridge, Centers for random walks on trees, *SIAM J. on Discrete Mathematics* **23** No. 1 (2009), 300-318.

[4] A. Beveridge and L. Lovász, Random walks and the regeneration time, *J. Graph Theory* **29** (1998), 57-62.

[5] D. Coppersmith, P. Tetali, and P. Winkler, Collisions among Random Walks on a Graph, *SIAM J. on Discrete Mathematics* **6** No. 3 (1993), 363–374.

[6] P. Doyle and J. L. Snell, Random walks and electric networks, (derived from their Carus Monograph, Mathematical Association of America, 1984), http://arxiv.org/abs/math/0001057.

[7] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, Cambridge (2002).

[8] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, Academic Press, Orlando, FL (1985).

[9] L. Lovász, Random Walks on Graphs: a Survey, *Combinatorics, Paul Erdős is Eighty, Vol. II*, (eds. D. Miklós, V. T. Sós and T. Szőnyi) J. Bolyai Math. Soc. (1996), 353–397.

[10] L. Lovász and P. Winkler, Efficient stopping rules for Markov chains, *Proceedings of the 27th ACM Symposium on the Theory of Computing* (1995), 76–82.

[11] L. Lovász and P. Winkler, Mixing of random walks and other diffusions on a graph, *Surveys in Combinatorics*, P. Rowlinson (editor), London Math. Soc. Lecture Note Series 218, pp. 119-154, 1995.

[12] L. Lovász and P. Winkler, Reversal of Markov Chains and the Forget Time, *Cominatorics, Probability and Computing* **7** (1998), 189–204.

[13] L. Lovász and P. Winkler, Mixing Times, *Microsurveys in Discrete Probability,* (ed. D. Aldous and J. Propp), DIMACS Series in Discrete Math. and Theor. Comp. Sci., AMS (1998), 85-133.

[14] J. W. Pitman, Occupation measures for Markov chains, *Adv. Appl. Prob.* **9** (1977), 69-86.