A Canonical Representation for Aggregated Markov Processes

BY BRET LARGET

Duquesne University

Abstract

A deterministic function of a Markov process is called an aggregated Markov process. We give necessary and sufficient conditions for the equivalence of continuous-time aggregated Markov processes. For both discrete- and continuous-time, we show that any aggregated Markov process which satisfies mild regularity conditions can be directly converted to a canonical representation which is unique for each class of equivalent models, and furthermore, is a minimal parameterization of all that can be identified about the underlying Markov process. Hidden Markov models on finite state spaces may be framed as aggregated Markov processes by expanding the state space and thus also have canonical representations.

The author completed the major part of this work as part of his Ph.D. dissertation at the University of California at Berkeley.

AMS 1991 subject classifications. Primary 60J27, Secondary 60J10

Key words and phrases. aggregated Markov process, canonical representation, hidden Markov model, identifiability, equivalence, minimal parameterization

1 Introduction Consider a continuous-time finite state space Markov process, $\{X(t), t \geq 0\}$ with state space $\mathcal{S} = \{1, 2, \ldots, n\}$, initial distribution $\boldsymbol{p}^T$, and stationary transition matrix $P(t) = e^{Qt}$, in which individual states have been collected together into several aggregates, partitioning the state space. Suppose that the state is not directly observable, but that a deterministic function of the underlying Markov process $Y(t) = f(X(t))$ which maps the state space $\mathcal{S}$ to the aggregate set $\mathcal{A} = \{\alpha_1, \ldots, \alpha_m\}$ can be observed. This observable stochastic process of aggregates $\{Y(t), t \geq 0\}$ is called an *aggregated Markov process* (AMP).

AMPs are a subclass of *hidden Markov models* (HMMs), which more generally allow the observed $Y(t)$ to be randomly determined given $X(t)$. When the HMM takes values on a finite set, expanding the state space to be the pairs $(X(t), Y(t))$ and defining f to be the projection function $f(x, y) = y$ allows this class of HMMs to be treated as AMPs on a finite state space.

Fredkin and Rice [4] and Kienker [8] are among several authors to study AMPs as a model for ion channel kinetics. Researchers in many fields apply HMMs to model sequences of dependent random variables. Some varied examples include molecular biology (see Baldi et al. [1] or Felsenstein and Churchill [3]), speech recognition (see Rabiner [9]), and climatology (see Zucchini and Guttorp [11]).

Typically, several states are mapped to the same aggregate, and in general, an AMP is no longer Markovian. Clearly, the aggregation of states can hide information about the underlying Markov process. A natural question to ask is, what about the distribution of the underlying Markov process can be identified from the distribution of the observed AMP?

One well-studied approach to this problem is to find conditions under which the finite-dimensional distributions of two representations are equal. Two such representations are called *equivalent*. Blackwell and Koopmans (1957) [2] first posed this question for stationary discrete-time AMPs. Thirty five years later, Ito et al. (1992) [7] finally gave a complete solution, restated in the notation of this paper in Proposition 1. We extend this chain of results to continuous-time AMPs in Theorem 1.

An alternative approach to the identifiability problem is to find a canonical representative for each equivalence class of equivalent representations. The main result of this paper is a solution to this problem under mild regularity conditions, contained in Theorem 2.

2 Equivalence of Continuous-time AMPs The probability of any observable sequence is the sum of the probabilities of all of the sequences of the underlying Markov process in which the states visited are in the proper aggregates. These sums may be expressed quite compactly with matrix notation. Furthermore, examining vector spaces which arise from these matrix expressions is the means through which the results on equivalence follow.

The *transition number* of a finite sequence will be one less than the length of the sequence and will be denoted by $|\boldsymbol{y}|$, so $|(y_0,\ldots,y_k)| = k$. Let $\Sigma(\mathcal{A})$ be the set of all finite sequences of aggregates and $\Sigma(\mathcal{T})$ be the set of all nondecreasing finite sequences of times whose first element is 0. A pair $(\boldsymbol{y},\boldsymbol{t})$ where $|\boldsymbol{y}| = |\boldsymbol{t}|$ is called an *observable sequence*. The set of all observable sequences is $\Sigma(\mathcal{A},\mathcal{T})$. For any observable sequence $(\boldsymbol{y},\boldsymbol{t})$ with $|\boldsymbol{y}| = k$, and square matrix M define the $(\boldsymbol{y},\boldsymbol{t})$*-power of the matrix* M to be

$$M^{(\boldsymbol{y},\boldsymbol{t})} = \begin{cases} I_{y_0}e^{M(t_1-t_0)}I_{y_1}e^{M(t_2-t_1)}I_{y_2}\cdots I_{y_{k-1}}e^{M(t_k-t_{k-1})}I_{y_k} & \text{for } k \geq 1 \\ I_{y_0} & \text{for } k = 0 \end{cases}$$

where I_α is the $n \times n$ diagonal matrix which is the identity on the submatrix where $f(i) = f(j) = \alpha$ and is zero elsewhere. Similarly, define the $\boldsymbol{y}$*-power of the matrix* M to be

$$M^{\boldsymbol{y}} = \begin{cases} I_{y_0}MI_{y_1}MI_{y_2}\cdots I_{y_{k-1}}MI_{y_k} & \text{for } k \geq 1 \\ I_{y_0} & \text{for } k = 0 \end{cases}$$

With this notation probabilities of observable sequences are compactly expressed as

$$\mathbb{P}\Big(Y(0) = y_0, Y(t_1) = y_1, \ldots, Y(t_k) = y_k\Big) = \boldsymbol{p}^T Q^{(\boldsymbol{y},\boldsymbol{t})}\mathbf{1}$$

where $\mathbf{1}$ is a column vector of ones. Vector spaces based on vectors of the form $\boldsymbol{p}^T Q^{(\boldsymbol{y},\boldsymbol{t})}$ and $Q^{(\boldsymbol{y},\boldsymbol{t})}\mathbf{1}$ are vital in what follows.

The standard *representation* for an AMP will be the quintuplet $(\mathcal{S},\boldsymbol{p}^T,Q,\mathbf{1},f)$, which fully determines its distribution. Later in the paper we will consider representations in which $\boldsymbol{p}^T$ need not be stochastic, Q need not be a proper generator, and $\mathbf{1}$ may be replaced by a different vector. Representations of discrete-time AMPs will be similar with the probability transition matrix P replacing the generator matrix Q. When discussing two AMPs, the representation of one will be distinguished by the use of tildes over the symbols.

Definition 1 *Two AMPs with representations* $(\mathcal{S}, \boldsymbol{p}^T, Q, \mathbf{1}, f)$ *and* $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{p}}^T, \tilde{Q}, \tilde{\mathbf{1}}, \tilde{f})$ *are defined to be equivalent if and only if*

$$\boldsymbol{p}^T Q^{(\boldsymbol{y},\boldsymbol{t})}\mathbf{1} = \tilde{\boldsymbol{p}}^T \tilde{Q}^{(\boldsymbol{y},\boldsymbol{t})}\tilde{\mathbf{1}} \qquad \textit{for all } (\boldsymbol{y},\boldsymbol{t}) \in \Sigma(\mathcal{A},\mathcal{T}) \tag{1}$$

The next lemma shows that (1) is equivalent to a simpler condition that does not involve matrix exponentials or time, and hence is a much easier condition to check for equivalence. The proof of the lemma is shown in the appendix.

Lemma 1 $(\mathcal{S}, \boldsymbol{p}^T, Q, \mathbf{1}, f)$ *and* $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{p}}^T, \tilde{Q}, \tilde{\mathbf{1}}, \tilde{f})$ *are equivalent if and only if*

$$\boldsymbol{p}^T Q^{\boldsymbol{y}}\mathbf{1} = \tilde{\boldsymbol{p}}^T \tilde{Q}^{\boldsymbol{y}}\mathbf{1} \qquad \textit{for all } \boldsymbol{y} \in \Sigma(\mathcal{A}) \tag{2}$$

Ito et al. define subspaces for discrete-time AMPs analogously to those in Definition 2, with $P^{\boldsymbol{y}}$ replacing $Q^{(\boldsymbol{y},\boldsymbol{t})}$. In a slight modification, we have replaced abstract quotient subspaces with orthogonal subspaces to remove the ambiguity in the subsequent definitions of some orthogonal projection matrices.

Definition 2 Each representation $(\mathcal{S}, \boldsymbol{p}^T, Q, \mathbf{1}, f)$ is associated with:

(a) *The state space,* $V = \mathbb{R}^n$

(b) *The reachable subspace,* $V(\boldsymbol{p}^T) = \mathrm{span}\left\{\boldsymbol{p}^T Q^{(\boldsymbol{y},\boldsymbol{t})} \;\middle|\; \text{for all } (\boldsymbol{y},\boldsymbol{t}) \in \Sigma(\mathcal{A},\mathcal{T})\right\}$

(c) *The null subspace,* $V^N = \left\{\boldsymbol{v}^T \in V \;\middle|\; \boldsymbol{v}^T Q^{(\boldsymbol{y},\boldsymbol{t})}\mathbf{1} = 0 \text{ for all } (\boldsymbol{y},\boldsymbol{t}) \in \Sigma(\mathcal{A},\mathcal{T})\right\}$

(d) *The imaginary subspace,* $W^I = \mathrm{span}\left\{(Q^{(\boldsymbol{y},\boldsymbol{t})}\mathbf{1})^T \;\middle|\; \text{for all } (\boldsymbol{y},\boldsymbol{t}) \in \Sigma(\mathcal{A},\mathcal{T})\right\}$

(e) *The effective subspace,* $W^E = V(\boldsymbol{p}^T) \cap W^I$

Note that the subspaces are comprised of row vectors.

The function f partitions the underlying states. We define classes of block diagonal matrices with respect to mappings f and $\tilde{f}$ to be

$$\begin{aligned}
\Delta(f) &= \left\{A \,\middle|\, A=[a_{ij}] \text{ is an } n\times n \text{ matrix where } f(i)\neq f(j) \text{ implies } a_{ij}=0\right\} \\
\Delta(\tilde{f}) &= \left\{A \,\middle|\, A=[a_{ij}] \text{ is an } \tilde{n}\times \tilde{n} \text{ matrix where } \tilde{f}(i)\neq \tilde{f}(j) \text{ implies } a_{ij}=0\right\} \\
\Delta(f,\tilde{f}) &= \left\{A \,\middle|\, A=[a_{ij}] \text{ is an } n\times \tilde{n} \text{ matrix where } f(i)\neq \tilde{f}(j) \text{ implies } a_{ij}=0\right\}
\end{aligned}$$

Proposition 1 is a restatement of Theorem 2 in [7], with modifications reflecting the changes in Definition 2 and inclusion of statements supressed in their proof.

Proposition 1 (Ito, Amari, and Kobayashi) *Two discrete-time AMPs with representations* $(\mathcal{S}, \boldsymbol{p}^T, P, \mathbf{1}, f)$ *and* $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{p}}^T, \tilde{P}, \tilde{\mathbf{1}}, \tilde{f})$ *are equivalent if and only if there exist* $A, B \in \Delta(f)$, $\tilde{A}, \tilde{B} \in \Delta(\tilde{f})$, *and* $S \in \Delta(f, \tilde{f})$ *such that*

(3) $$\text{(a) } APS = S\tilde{P}\tilde{A}, \quad \text{(b) } \boldsymbol{p}^T S = \tilde{\boldsymbol{p}}^T \tilde{A}, \quad \text{(c) } A\mathbf{1} = S\tilde{\mathbf{1}}$$

(4) $$\text{(a) } PB = BPB, \quad \text{(b) } APB = APA, \quad \text{(c) } \boldsymbol{p}^T B = \boldsymbol{p}^T A, \quad \text{(d) } B\mathbf{1} = \mathbf{1}$$

(5) $$\text{(a) } \tilde{P}\tilde{B} = \tilde{B}\tilde{P}\tilde{B}, \quad \text{(b) } \tilde{A}\tilde{P}\tilde{B} = \tilde{A}\tilde{P}\tilde{A}, \quad \text{(c) } \tilde{\boldsymbol{p}}^T \tilde{B} = \tilde{\boldsymbol{p}}^T \tilde{A}, \quad \text{(d) } \tilde{B}\tilde{\mathbf{1}} = \tilde{\mathbf{1}}$$

Furthermore, W^E *and* $\tilde{W}^E$ *are of equal dimension. In addition,* A, $\tilde{A}$, B, *and* $\tilde{B}$ *may be chosen to be the orthogonal projection matrices into* W^E, $\tilde{W}^E$, W^I, *and* $\tilde{W}^I$ *respectively.* S *may be chosen to satisfy the equations:*

(6) $$(\boldsymbol{p}^T P^{\boldsymbol{y}} A)S = (\tilde{\boldsymbol{p}}^T \tilde{P}^{\boldsymbol{y}} \tilde{A}) \qquad \text{for all } \boldsymbol{y} \in \Sigma(\mathcal{A})$$

and

(7) $$\boldsymbol{v}^T S = 0 \qquad \text{for all } \boldsymbol{v}^T \in (W^E)^\perp$$

S represents an invertible mapping between the two effective subspaces. As a matrix, S need not be square, much less be invertible.

To extend this result to continuous-time, the AMPs are discretized by constructing a discrete probability transition matrix by first-order approximation of $P(t)$ for small t, and applying Proposition 1 to the discretized AMPs. This is related to the construction of continuous-time Markov processes by uniformization. The various subspaces of a continuous-time AMP turn out to be the same as the corresponding subspaces of its discretized AMP, regardless of the choice of t. Exploiting the interplay between the matrices and the vector subspaces gives this theorem, whose proof is given in Section 5.

Theorem 1 *Let $(\mathcal{S}, \boldsymbol{p}^T, Q, \mathbf{1}, f)$ and $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{p}}^T, \tilde{Q}, \tilde{\mathbf{1}}, \tilde{f})$ be the representations of two continuous-time AMPs, with A and $\tilde{A}$ the orthogonal projection matrices into W^E and $\tilde{W}^E$ respectively. Then the representations are equivalent if and only if there exists a matrix $S \in \Delta(f, \tilde{f})$ such that*

$$(a)\ (AQA)S = S(\tilde{A}\tilde{Q}\tilde{A}), \quad (b)\ (\boldsymbol{p}^T A)S = \tilde{\boldsymbol{p}}^T \tilde{A}, \quad (c)\ A\mathbf{1} = S(\tilde{A}\tilde{\mathbf{1}}) \tag{8}$$

Thinking of the matrices AQA and $\tilde{A}\tilde{Q}\tilde{A}$ as *effective generators*, $\boldsymbol{p}^T A$ and $\tilde{\boldsymbol{p}}^T \tilde{A}$ as *effective initial distributions*, and $A\mathbf{1}$ and $\tilde{A}\tilde{\mathbf{1}}$ as *effective summing vectors*, the theorem says that equivalence occurs if and only if there exists a block diagonal similarity relationship between the effective generators for which the effective initial distributions and summing vectors correspond.

3 A Canonical Representation for AMPs The conditions in the previous theorem involve orthogonal projection matrices into the effective subspaces. Ito et al. show that the subspaces are generated by vectors from observable sequences with transition number less than n. However, there may still be a considerable amount of work involved in determining equivalence.

Under mild assumptions, Theorem 2 gives conditions for equivalence which may be checked in a single direct computation by converting the standard representation to a canonical representation. Furthermore, this canonical representation contains a minimal parameterization of all identifiable information for the class of equivalent AMPs satisfying the assumptions.

Since submatrices induced by the partitioning into aggregates will be of interest, without loss of generality, order all states so that states mapped to the same aggregate are numbered consecutively and the aggregates appear in a prespecified order. Let n_α be the number of states in aggregate α for $\alpha \in \mathcal{A}$ and note that $n = \sum n_\alpha$. Define $\boldsymbol{p}_\alpha^T$ to be the generic subvector of $\boldsymbol{p}^T$ for which $f(i) = \alpha$ and $Q_{\alpha\beta}$ to be the generic $n_\alpha \times n_\beta$ matrix block of Q for which $f(i) = \alpha$ and $f(j) = \beta$. Terms with tildes are defined in an analogous fashion.

A notation for Krylov sequences and matrices must be introduced to state the canonical representation. For a vector $\boldsymbol{x}$ and a square matrix M, a *Krylov sequence* is given by $\boldsymbol{x}, M\boldsymbol{x}, M^2\boldsymbol{x}, \ldots$. Similarly, for row vectors $\boldsymbol{x}^T, \boldsymbol{x}^T M, \boldsymbol{x}^T M^2, \ldots$ is a Krylov sequence. A matrix is a *Krylov matrix* if its columns (or rows) are the initial part of a Krylov sequence. Let $K(\boldsymbol{x}, M, n)$ be the matrix whose jth column is $M^{j-1}\boldsymbol{x}$ for $1 \leq j \leq n$ and $K(\boldsymbol{x}^T, M, n)$ be the matrix whose ith row is $\boldsymbol{x}^T M^{i-1}$ for $1 \leq i \leq n$.

Let $\bar{\boldsymbol{p}}^T$ be the equilibrium distribution of the AMP and consider the following list of regularity conditions.

C1. For each α, there exist n_α observable sequences $\{(\boldsymbol{y}, \boldsymbol{t})_i\}$ such that the vectors $\{\boldsymbol{p}^T Q^{(\boldsymbol{y},\boldsymbol{t})_i} I_\alpha\}$ are linearly independent.

C2. For each α, there exist n_α observable sequences $\{(\boldsymbol{y}, \boldsymbol{t})_i\}$ such that the vectors $\{I_\alpha Q^{(\boldsymbol{y},\boldsymbol{t})_i} \mathbf{1}\}$ are linearly independent.

C3. The eigenvalues of $Q_{\alpha\alpha}$ are distinct and not equal to 0 for each $\alpha \in \Sigma(\mathcal{A})$.

C4. $\bar{\boldsymbol{p}}_\alpha^T \boldsymbol{u}_\alpha \neq 0$ for every right eigenvector $\boldsymbol{u}_\alpha$ of $Q_{\alpha\alpha}$ for each $\alpha \in \Sigma(\mathcal{A})$.

C5. $\boldsymbol{v}_\alpha^T \mathbf{1} \neq 0$ for every left eigenvector $\boldsymbol{v}_\alpha^T$ of $Q_{\alpha\alpha}$ for each $\alpha \in \Sigma(\mathcal{A})$.

C6. $K(\boldsymbol{p}^T, Q_{\alpha\alpha}, n_\alpha)$ is invertible for each $\alpha \in \Sigma(\mathcal{A})$.

C7. $K(\mathbf{1}_\alpha, Q_{\alpha\alpha}, n_\alpha)$ is invertible for each $\alpha \in \Sigma(\mathcal{A})$.

In the many papers on AMPs, various authors have assumed different regularity conditions. Together, conditions *C1* and *C2* are equivalent to Gilbert's [6] original regularity conditions, adapted to continuous-time, and are also equivalent to the dimension of the effective space being n. Fredkin and Rice [5] assume conditions *C3–C5* in showing that the two-dimensional dwell-time distributions determine equivalence. We assume *C6–C7* in construction of the canonical representation and claim that these assumptions are weaker than *C3–C5* when $\boldsymbol{p}^T$ is the equilibrium distribution but stronger than *C1–C2* for general $\boldsymbol{p}^T$. The second claim is proven in Section 6.

The next theorem is the major result of the paper. Under mild regularity conditions, a canonical representation can be directly computed from any given representation which will be the same for all equivalent representations satisfying the same regularity conditions.

Theorem 2 *Suppose that a continuous-time AMP represented by* $(\mathcal{S}, \boldsymbol{p}^T, Q, \mathbf{1}, f)$ *satisfies conditions C6 and C7. Define* $C \in \Delta(f)$ *as the block diagonal matrix with* $C_\alpha = K(\mathbf{1}_\alpha, Q_{\alpha\alpha}, n_\alpha)$. *Let* $\boldsymbol{a}^T = \boldsymbol{p}^T C$, $H = C^{-1}QC$, *and* $\boldsymbol{b} = C^{-1}\mathbf{1}$. *Then*

$$\mathbb{P}\Big(Y(0) = y_0, Y(t_1) = y_1, \ldots, Y(t_k) = y_k\Big) = \boldsymbol{a}^T H^{(\boldsymbol{y},\boldsymbol{t})}\boldsymbol{b} \quad \textit{for all } (\boldsymbol{y}, \boldsymbol{t}) \in \Sigma(\mathcal{A}, \mathcal{T})$$

Furthermore, for any other equivalent AMP represented by $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{p}}^T, \tilde{Q}, \tilde{\mathbf{1}}, \tilde{f})$ *which satisfies conditions C6 and C7, for* $\tilde{C}$ *defined analogously to* C,

$$\boldsymbol{p}^T C = \tilde{\boldsymbol{p}}^T \tilde{C}, \quad C^{-1}QC = \tilde{C}^{-1}\tilde{Q}C, \quad \textit{and} \quad C^{-1}\mathbf{1} = \tilde{C}^{-1}\tilde{\mathbf{1}} \tag{9}$$

The representation

$$(\mathcal{S}, \boldsymbol{a}^T, H, \boldsymbol{b}, f) = (\mathcal{S},\, \boldsymbol{p}^T C,\, C^{-1}QC,\, C^{-1}\mathbf{1},\, f)$$

is called the canonical representation of the AMP.

Before proving the theorem in Section 6, we make some observations on the form of the canonical representation.

(a) Except for the trivial case in which each state is mapped to its own aggregate, the canonical representation will not be standard.

(b) $H_{\alpha\alpha}$, the diagonal α-block of H, is a companion matrix of $Q_{\alpha\alpha}$, meaning that the elements of the first subdiagonal of $H_{\alpha\alpha}$ are all 1's, the rightmost column of $H_{\alpha\alpha}$ consists of the negative coefficients of the characteristic equation of $Q_{\alpha\alpha}$, and the rest of the block is zero.

(c) The vector $\boldsymbol{b}$ contains exactly m ones, one in the first row corresponding to each aggregate, with the remaining elements being 0. In addition, $\boldsymbol{b}$ is a right eigenvector of H with eigenvalue 0.

The discrete-time analogue of Theorem 2 is true if the probability transition matrix P is substituted for Q in conditions *C6* and *C7*, and in the statement of the theorem. The only difference in the canonical representation is that $\boldsymbol{b}$ is a right eigenvector with eigenvalue 1 instead of 0. The proof is perfectly analogous to the proof of Theorem 2 and is therefore omitted.

4 A Minimal Parameterization The canonical representation is the same for all equivalent AMPs which satisfy regularity conditions *C6–C7*. Furthermore, this parameterization determines the probability of observing any observable sequence. Together, these two statements mean that the free parameters in the canonical representation form a minimal parameterization of everything which may be identified from the AMP. A natural question to ask is, how many parameters are necessary to completely specify the AMP?

Notice that each diagonal block in the canonical representation contains n_α free parameters, the coefficients of the characteristic equation. The off-diagonal blocks are free, up to the constraint that the vector $\boldsymbol{b}$ is a right eigenvector of H with eigenvalue 0. If the AMP is at equilibrium, the initial distribution is determined by H. The number of free parameters is

$$\sum_{i=1}^{m} n_i + \sum_{1\le i<j\le m} 2n_in_j - n = 2\sum_{1\le i<j\le m} n_in_j \tag{10}$$

An arbitrary initial distribution is allowed with $n-1$ more parameters.

Assuming the rank of $Q_{\alpha\beta}$ is full for all off-diagonal matrix blocks, this total agrees with the number of free parameters found by Fredkin and Rice [5] for AMPs at equilibrium.

For a continuous-time AMP which begins at equilibrium and satisfies conditions *C6* and *C7*, the free parameters of the canonical generator H are a minimal parameterization of what can be identified from the underlying Markov process.

Most objects of interest can be computed directly from the canonical representation. Theorem 2 gives the expression for the probability of any observable sequence. As another example, the two-dimensional equilibrium dwell-time distributions from Fredkin and Rice [5] can be computed by

$$\begin{aligned} f_{\alpha\beta}(s,t) &= \left(\frac{\boldsymbol{p}^T Q_{\alpha\alpha}}{\boldsymbol{p}^T Q_{\alpha\alpha}\mathbf{1}}\right) e^{Q_{\alpha\alpha}s} Q_{\alpha\beta} e^{Q_{\beta\beta}t}(-Q_{\beta\beta}\mathbf{1}_\beta) \\ &= \left(\frac{\boldsymbol{a}^T H_{\alpha\alpha}}{\boldsymbol{a}^T H_{\alpha\alpha}\boldsymbol{b}}\right) e^{H_{\alpha\alpha}s} H_{\alpha\beta} e^{H_{\beta\beta}t}(-H_{\beta\beta}\boldsymbol{b}_\beta) \end{aligned}$$

5 Proof of Theorem 1 We begin by showing that the reachable subspace and null subspace defined in Definition 2 have equivalent characterizations based on simpler vectors.

Lemma 2 *The reachable subspace can be represented in these equivalent ways,*

(i) $\mathcal{R}_1 = \operatorname{span}\left\{\boldsymbol{p}^T Q^{(\boldsymbol{y},\boldsymbol{t})} \,\middle|\, \text{for all } (\boldsymbol{y},\boldsymbol{t}) \in \Sigma(\mathcal{A},\mathcal{T})\right\}$

(ii) $\mathcal{R}_2 = \operatorname{span}\left\{\boldsymbol{p}^T Q^{\boldsymbol{y}} \,\middle|\, \text{for all } \boldsymbol{y} \in \Sigma(\mathcal{A})\right\}$

(iii) $\mathcal{R}_3 = \operatorname{span}\left\{\boldsymbol{p}^T (I+Qt)^{\boldsymbol{y}} \,\middle|\, \text{for all } \boldsymbol{y} \in \Sigma(\mathcal{A}) \text{ with } t>0 \text{ fixed}\right\}$

and the null subspace can be represented in these equivalent ways.

(i) $\mathcal{N}_1 = \left\{\boldsymbol{v}^T \,\middle|\, \boldsymbol{v}^T Q^{(\boldsymbol{y},\boldsymbol{t})}\mathbf{1} = 0 \text{ for all } (\boldsymbol{y},\boldsymbol{t}) \in \Sigma(\mathcal{A},\mathcal{T})\right\}$

(ii) $\mathcal{N}_2 = \left\{\boldsymbol{v}^T \,\middle|\, \boldsymbol{v}^T Q^{\boldsymbol{y}}\mathbf{1} = 0 \text{ for all } \boldsymbol{y} \in \Sigma(\mathcal{A})\right\}$

(iii) $\mathcal{N}_3 = \left\{\boldsymbol{v}^T \,\middle|\, \boldsymbol{v}^T (I+Qt)^{\boldsymbol{y}}\mathbf{1} = 0 \text{ for all } \boldsymbol{y} \in \Sigma(\mathcal{A}) \text{ with } t>0 \text{ fixed}\right\}$

Proof: ($\mathcal{R}_1 \subset \mathcal{R}_2$) Fix $(\boldsymbol{y},\boldsymbol{t}) \in \Sigma(\mathcal{A},\mathcal{T})$ letting $k = |\boldsymbol{y}|$ and $u_i = t_i - t_{i-1}$ for $1 \le i \le k$. Then

$$\boldsymbol{p}^T Q^{(\boldsymbol{y},\boldsymbol{t})} = \boldsymbol{p}^T I_{y_0} e^{Qu_1} \cdots I_{y_{k-1}} e^{Qu_k} I_{y_k}$$

It is well known that the matrix exponential of an $n \times n$ matrix Mu can be written as a finite linear combination of powers of M.

$$e^{Mu} = \sum_{i=0}^{n-1} c_i(u) M^i$$

The relationship between the ordinary kth power of a matrix M and the $\boldsymbol{y}$-powers of the matrix for all sequences with transition number k is easily determined using $I = \sum I_\alpha$.

$$M^k = \sum_{|\boldsymbol{y}|=k} M^{\boldsymbol{y}} \tag{11}$$

From these two relationships we see that $\boldsymbol{p}^T Q^{(\boldsymbol{y},\boldsymbol{t})}$ is a finite linear combination of vectors of the form $\boldsymbol{p}^T Q^{\boldsymbol{z}}$, and hence is in $\mathcal{R}_2$.

($\mathcal{R}_2 \subset \mathcal{R}_1$) Use induction on k. Suppose that $|\boldsymbol{y}| = 0$ so that $Q^{\boldsymbol{y}} = I_\alpha$ for some $\alpha \in \mathcal{A}$. Then $\boldsymbol{p}^T Q^{\boldsymbol{y}} = \boldsymbol{p}^T I_\alpha \in \mathcal{R}_1$. Assume that $\boldsymbol{p}^T Q^{\boldsymbol{y}} \in \mathcal{R}_1$ for all $|\boldsymbol{y}| \leq k$. For $|\boldsymbol{y}| = k+1$, let $\boldsymbol{y} = (y_0, y_1, \ldots, y_{k+1})$, and $\boldsymbol{z} = (y_0, \ldots, y_k)$. Since the derivative of e^{Qt} is Qe^{Qt},

$$\begin{aligned}
\boldsymbol{p}^T Q^{\boldsymbol{y}} &= \boldsymbol{p}^T Q^{\boldsymbol{z}} Q I_{y_{k+1}} \\
&= \frac{\mathrm{d}}{\mathrm{dt}} \left[\boldsymbol{p}^T Q^{\boldsymbol{z}} e^{Qt} I_{y_{k+1}}\right]_{t=0} \\
&= \lim_{t \downarrow 0} \frac{\boldsymbol{p}^T Q^{\boldsymbol{z}} e^{Qt} I_{y_{k+1}} - \boldsymbol{p}^T Q^{\boldsymbol{z}} I_{y_{k+1}}}{t} \in \mathcal{R}_1
\end{aligned}$$

since $\boldsymbol{p}^T Q^{\boldsymbol{z}} \in \mathcal{R}_1$ by assumption and $\mathcal{R}_1$ is easily seen to be closed under right multiplication by e^{Qt} for all $t \geq 0$ and I_α for all $\alpha \in \Sigma(\mathcal{A})$.

That $\mathcal{R}_2 = \mathcal{R}_3$ follows from a straightforward use of induction on $|\boldsymbol{y}|$ in both directions.

An identical argument can be used to show that analogously defined characterizations of the imaginary subspace exist, and the results on the null subspace follow directly since $V^N = (W^I)^\perp$.

□

Lemma 3 mimics Proposition 1 where the generator matrix plays the role of the probability transition matrix.

Lemma 3 *Two continuous-time AMPs with representations* $(\mathcal{S}, \boldsymbol{p}^T, Q, \mathbf{1}, f)$ *and* $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{p}}^T, \tilde{Q}, \tilde{\mathbf{1}}, \tilde{f})$ *are equivalent if and only if there exist* $A, B \in \Delta(f)$, $\tilde{A}, \tilde{B} \in \Delta(\tilde{f})$, *and* $S \in \Delta(f, \tilde{f})$ *such that*

$$\text{(a) } AQS = S\tilde{Q}\tilde{A}, \quad \text{(b) } \boldsymbol{p}^T S = \tilde{\boldsymbol{p}}^T \tilde{A}, \quad \text{(c) } A\mathbf{1} = S\tilde{\mathbf{1}} \tag{12}$$

$$\text{(a) } QB = BQB, \quad \text{(b) } AQB = AQA, \quad \text{(c) } \boldsymbol{p}^T B = \boldsymbol{p}^T A, \quad \text{(d) } B\mathbf{1} = \mathbf{1} \tag{13}$$

$$\text{(a) } \tilde{Q}\tilde{B} = \tilde{B}\tilde{Q}\tilde{B}, \quad \text{(b) } \tilde{A}\tilde{Q}\tilde{B} = \tilde{A}\tilde{Q}\tilde{A}, \quad \text{(c) } \tilde{\boldsymbol{p}}^T \tilde{B} = \tilde{\boldsymbol{p}}^T \tilde{A}, \quad \text{(d) } \tilde{B}\tilde{\mathbf{1}} = \tilde{\mathbf{1}} \tag{14}$$

Proof: In the proof of necessity, the equivalent continuous-time AMPs are discretized by forming probability transition matrices as first order approximations to $P(t)$ for $t > 0$ small enough to ensure that the matrices $P_t = I + tQ$ and $\tilde{P}_t = \tilde{I} + t\tilde{Q}$ are stochastic. As shown by Rýden [10],

$$\boldsymbol{p}^T Q^{(\boldsymbol{y},\boldsymbol{t})}\mathbf{1} = \tilde{\boldsymbol{p}}^T \tilde{Q}^{(\boldsymbol{y},\boldsymbol{t})}\tilde{\mathbf{1}} \qquad \text{for all } (\boldsymbol{y},\boldsymbol{t}) \in \Sigma(\mathcal{A},\mathcal{T})$$

implies that

$$\boldsymbol{p}^T (I + tQ)^{\boldsymbol{y}}\mathbf{1} = \tilde{\boldsymbol{p}}^T (\tilde{I} + t\tilde{Q})^{\boldsymbol{y}}\tilde{\mathbf{1}} \qquad \text{for all } \boldsymbol{y} \in \Sigma(\mathcal{A})$$

by Taylor series expansion of the matrix exponentials, equating of coefficients, and straightforward algebraic manipulation. Thus the discrete-time AMPs with representations $(\mathcal{S}, \boldsymbol{p}^T, I + tQ, f)$ and $(\mathcal{S}, \tilde{\boldsymbol{p}}^T, \tilde{I} + t\tilde{Q}, \tilde{f})$ are also equivalent.

Proposition 1 implies the existence of A_t, $\tilde{A}_t$, B_t, and $\tilde{B}_t$, orthogonal projection matrices into subspaces W_t^E, $\tilde{W}_t^E$, W_t^I, and $\tilde{W}_t^I$ defined for the discrete processes respectively, and matrix S_t, which maps corresponding elements of W_t^E to $\tilde{W}_t^E$, all of which satisfy (3) – (5), where $P = I + tQ$ and $\tilde{P} = \tilde{I} + t\tilde{Q}$. By Lemma 2, these matrices are independent of the choice of t and have the same interpretations for the corresponding subspaces defined for the continuous processes, so we drop the subscripts from the notation. Note also that (6) and (7) imply that

$$AS = S\tilde{A} = S \tag{15}$$

Since $W^E \subset W^I$ and A and B are orthogonal projection matrices, $AB = BA = A$, $A^2 = A$, and $B^2 = B$. Analogous statements hold for $\tilde{A}$ and $\tilde{B}$. With this in mind, (12) – (14) follow from (3) – (5) after some algebra. As an example, consider (3a).

$$\begin{aligned} A(I + Qt)S = S(\tilde{I} + \tilde{Q}t)\tilde{A} \quad &\Rightarrow \quad t(AQS - S\tilde{Q}\tilde{A}) = S\tilde{A} - AS = 0 \\ &\Rightarrow \quad AQS = S\tilde{Q}\tilde{A} \end{aligned}$$

The other ten equations are derived in a straightforward fashion, and the proof of necessity is complete.

The proof of sufficiency is completed by direct calculation.

$$
\begin{aligned}
\boldsymbol{p}^T Q^{\boldsymbol{y}} \mathbf{1} &= \boldsymbol{p}^T Q^{\boldsymbol{y}} B \mathbf{1} && \text{by (13d)} \\
&= \boldsymbol{p}^T B (QB)^{\boldsymbol{y}} \mathbf{1} && \text{by (13a) and } I_\alpha B = B I_\alpha \\
&= \boldsymbol{p}^T A (QB)^{\boldsymbol{y}} \mathbf{1} && \text{by (13c)} \\
&= \boldsymbol{p}^T (AQ)^{\boldsymbol{y}} A \mathbf{1} && \text{by (13b) and } A I_\alpha = I_\alpha A \\
&= \boldsymbol{p}^T (AQ)^{\boldsymbol{y}} S \tilde{\mathbf{1}} && \text{by (12c)} \\
&= \boldsymbol{p}^T S (\tilde{Q}\tilde{A})^{\boldsymbol{y}} \tilde{\mathbf{1}} && \text{by (12a) and } I_\alpha S = S \tilde{I}_\alpha \\
&= \tilde{\boldsymbol{p}}^T \tilde{A} (\tilde{Q}\tilde{A})^{\boldsymbol{y}} \tilde{\mathbf{1}} && \text{by (12b)} \\
&= \tilde{\boldsymbol{p}}^T (\tilde{A}\tilde{Q})^{\boldsymbol{y}} \tilde{A} \tilde{\mathbf{1}} && \text{since } \tilde{A} I_\alpha = I_\alpha \tilde{A} \\
&= \tilde{\boldsymbol{p}}^T \tilde{A} (\tilde{Q}\tilde{B})^{\boldsymbol{y}} \tilde{\mathbf{1}} && \text{by (14b) and } \tilde{I}_\alpha \tilde{A} = \tilde{A} \tilde{I}_\alpha \\
&= \tilde{\boldsymbol{p}}^T \tilde{B} (\tilde{Q}\tilde{B})^{\boldsymbol{y}} \tilde{\mathbf{1}} && \text{by (14c)} \\
&= \tilde{\boldsymbol{p}}^T \tilde{Q}^{\boldsymbol{y}} \tilde{B} \tilde{\mathbf{1}} && \text{by (14a) and } \tilde{B} \tilde{I}_\alpha = \tilde{I}_\alpha \tilde{B} \\
&= \tilde{\boldsymbol{p}}^T \tilde{Q}^{\boldsymbol{y}} \tilde{\mathbf{1}} && \text{by (14d)}
\end{aligned}
$$

□

We are now in position to complete the proof.

Proof of Theorem 1: The theorem follows directly from Lemma 3. First, notice from Proposition 1 that it is possible to define A, $\tilde{A}$, B, and $\tilde{B}$ as orthogonal projection matrices into subspaces W_t^E, $\tilde{W}_t^E$, W_t^I, and $\tilde{W}_t^I$ respectively and that these definitions do not depend on the equivalence of the two AMPs. It follows that (13) and (14) hold when A, $\tilde{A}$, B, and $\tilde{B}$ are so defined, since each of these equations depends only on one AMP. Hence, two AMPs are equivalent if and only if (12) holds for A and $\tilde{A}$ defined as orthogonal projection matrices into W^E and $\tilde{W}^E$ respectively for an $S \in \Delta(f, \tilde{f})$. (8) follows from (12) and (15). □

6 Proof of Theorem 2

Under the regularity conditions we assume, the effective subspaces are full rank and the orthogonal projection matrices into these subspaces are the identity.

Lemma 4 *Regularity conditions* $\{C6–C7\}$ *imply* $\{C1–C2\}$.

Proof: Notice that *C1* is equivalent to $\dim V(\boldsymbol{p}^T) = n$ and that *C2* is equivalent to $\dim W^I = n$. For each α, consider the n_α aggregate sequences $\{\boldsymbol{y}_i\}$ where $\boldsymbol{y}_i$ is the sequence which repeats α i times. Condition *C6* implies that the set of vectors $\boldsymbol{p}^T Q^{\boldsymbol{y}_i}$ is linearly independent, since the α-blocks of these vectors are the rows of $K(\boldsymbol{p}^T, Q_{\alpha\alpha}, \mathbf{1}_\alpha)$. The collection of these vectors for all α is a linearly independent set of n vectors in characterization $\mathcal{R}_2$ of the reachable subspace in Lemma 2, so *C6* implies *C1*. A similar argument shows that *C7* implies *C2*. $\square$

By Lemma 4, *C6* – *C7* imply *C1* – *C2* which implies that $W^E = \tilde{W}^E = \mathbb{R}^n$ and orthogonal projection matrices A and $\tilde{A}$ are the identity. Theorem 1 reduces to equivalence between $(\mathcal{S}, \boldsymbol{p}^T, Q, \mathbf{1}, f)$ and $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{p}}^T, \tilde{Q}, \tilde{\mathbf{1}}, \tilde{f})$ if and only if there exists a block diagonal matrix S that satisfies

$$\text{(a) } QS = S\tilde{Q}, \quad \text{(b) } \boldsymbol{p}^T S = \tilde{\boldsymbol{p}}^T, \quad \text{(c) } S\mathbf{1} = \mathbf{1} \tag{16}$$

We now conclude the proof.

Proof of Theorem 2: The first assertion of Theorem 2 follows by simple matrix multiplication. Define $C \in \Delta(f)$ as the block diagonal matrix with $C_\alpha = K(\mathbf{1}_\alpha, Q_{\alpha\alpha}, n_\alpha)$. Let $\boldsymbol{a}^T = \boldsymbol{p}^T C$, $H = C^{-1} Q C$, and $\boldsymbol{b} = C^{-1}\mathbf{1}$. Condition *C6* implies that C is invertible. It follows immediately that for any $\boldsymbol{y} \in \mathcal{A}$,

$$\boldsymbol{p}^T Q^{\boldsymbol{y}} \mathbf{1} = (\boldsymbol{p}^T C)(C^{-1} Q C)^{\boldsymbol{y}} (C^{-1}\mathbf{1}) = \boldsymbol{a}^T H^{\boldsymbol{y}} \boldsymbol{b} \tag{17}$$

since $C I_\alpha = I_\alpha C$ for any $\alpha \in \mathcal{A}$. By Lemma 1, the canonical representation is equivalent to $(\mathcal{S}, \boldsymbol{p}^T, Q, \mathbf{1}, f)$.

The second assertion is that this canonical representation is the same for all equivalent AMPs, which we now show.

Conditions *C6* – *C7* state that the matrices $R_\alpha = K(\boldsymbol{p}_\alpha^T, Q_{\alpha\alpha}, n_\alpha)$ and $C_\alpha = K(\mathbf{1}_\alpha, Q_{\alpha\alpha}, n_\alpha)$ are invertible for each aggregate $\alpha \in \mathcal{A}$. $\tilde{R}_\alpha$ and $\tilde{C}_\alpha$ are defined similarly and are likewise invertible. Let R, C, $\tilde{R}$, and $\tilde{C}$ be the block diagonal matrices in $\Delta(f)$ with respective diagonal blocks R_α, C_α, $\tilde{R}_\alpha$, and $\tilde{C}_\alpha$.

By equivalence of the AMPs, the invertible block-diagonal matrix S satisfying (16) maps vectors of the form $\boldsymbol{p}^T Q^{\boldsymbol{y}}$ to $\tilde{\boldsymbol{p}}^T \tilde{Q}^{\boldsymbol{y}}$. Since the rows of R and $\tilde{R}$ are in this form, $RS = \tilde{R}$ which implies that $S = R^{-1}\tilde{R}$.

The ij element of the matrix $R_\alpha C_\alpha$ is $\boldsymbol{p}^T Q_{\alpha\alpha}^{i+j-2}\mathbf{1}_\alpha$ which equals $\tilde{\boldsymbol{p}}^T \tilde{Q}_{\alpha\alpha}^{i+j-2}\mathbf{1}_\alpha$ by Lemma 1 so that $RC = \tilde{R}\tilde{C}$, from which it follows that $C\tilde{C}^{-1} = R^{-1}\tilde{R} = S$. Substituting $C\tilde{C}^{-1}$ for S in (16) and matrix multiplication completes the proof. $\square$

7 A Numerical Example To make these ideas more concrete, we demonstrate them with an example. Suppose two underlying Markov processes on five states start at their equilibrium distributions $\boldsymbol{p}^T = (0.2, 0.2, 0.2, 0.2, 0.2)$ and $\tilde{\boldsymbol{p}}^T = (8/45, 2/9, 1/5, 8/45, 2/9)$ with generators

$$Q = \left[\begin{array}{rr|rrr} -9 & 6 & 3 & 0 & 0 \\ 6 & -12 & 0 & 6 & 0 \\ \hline 3 & 0 & -6 & 3 & 0 \\ 0 & 6 & 3 & -12 & 3 \\ 0 & 0 & 0 & 3 & -3 \end{array}\right] \quad \tilde{Q} = \left[\begin{array}{rr|rrr} -38/3 & 20/3 & 0 & 6 & 0 \\ 151/30 & -25/3 & 0 & 3/10 & 3 \\ \hline 0 & 0 & -3 & 3 & 0 \\ 6 & 0 & 3 & -37/3 & 10/3 \\ 3/10 & 3 & 3/10 & 31/15 & -17/3 \end{array}\right]$$

Let $f(1) = f(2) = \alpha_1$ and $f(3) = f(4) = f(5) = \alpha_2$. The corresponding matrices with Krylov matrix diagonal blocks are

$$C = \left[\begin{array}{rr|rrr} 1 & -3 & 0 & 0 & 0 \\ 1 & -6 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & -3 & 0 \\ 0 & 0 & 1 & -6 & 63 \\ 0 & 0 & 1 & 0 & -18 \end{array}\right] \quad \tilde{C} = \left[\begin{array}{rr|rrr} 1 & -6.0 & 0 & 0 & 0 \\ 1 & -3.3 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & -18.0 \\ 0 & 0 & 1 & -6.0 & 63.0 \\ 0 & 0 & 1 & -3.3 & 6.3 \end{array}\right]$$

C and $\tilde{C}$ are invertible, and condition *C7* also holds for both AMPs, so $W^E = \tilde{W}^E = V$ and $A = \tilde{A} = I$. The matrix

$$S = \left[\begin{array}{cc|ccc} 0 & 1 & 0 & 0 & 0 \\ 0.9 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.9 & 0.1 & 0 \end{array}\right]$$

satisfies Theorem 1 and the two AMPs are equivalent.

Following Theorem 2, $\boldsymbol{a}^T = \boldsymbol{p}^T C = \tilde{\boldsymbol{p}}^T \tilde{C} = (0.4, -1.8, 0.6, -1.8, 9)$, $\boldsymbol{b} = C^{-1}\mathbf{1} = \tilde{C}^{-1}\mathbf{1} = (1, 0, 1, 0, 0)^T$, and the canonical generator given by $H = C^{-1}QC = \tilde{C}^{-1}\tilde{Q}\tilde{C}$ is

$$H = \left[\begin{array}{cc|ccc} 0 & -72.0 & 0 & 18 & -378 \\ 1 & -21.0 & -1 & 9 & -126 \\ \hline 0 & -7.2 & 0 & 0 & -135 \\ -1 & 0.6 & 1 & 0 & -108 \\ 0 & -0.4 & 0 & 1 & -21 \end{array}\right]$$

Notice that each diagonal block is in companion form and that the first and third columns sum to the zero vector. There are $2 \times 2 \times 3 = 12$ free parameters in H, which is a minimal parameterization of all that is identifiable.

8 Acknowledgements I thank my dissertation advisor John Rice for his immense and varied help in the development of this article.

9 Appendix We prove Lemma 1.

Proof of Lemma 1: First assume that $(\mathcal{S}, \boldsymbol{p}^T, Q, \mathbf{1}, f)$ and $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{p}}^T, \tilde{Q}, \tilde{\mathbf{1}}, \tilde{f})$ are equivalent. For any $\boldsymbol{t} \in \mathcal{T}$ let $u_i = t_i - t_{i-1}$. Then by definition,

$$\boldsymbol{p}^T I_{y_0} e^{Qu_1} I_{y_1} \cdots I_{y_{k-1}} e^{Qu_k} I_{y_k} \mathbf{1} = \tilde{\boldsymbol{p}}^T \tilde{I}_{y_0} e^{\tilde{Q}u_1} \tilde{I}_{y_1} \cdots \tilde{I}_{y_{k-1}} e^{\tilde{Q}u_k} \tilde{I}_{y_k} \tilde{\mathbf{1}} \tag{18}$$

for all $k \geq 0$, $u_1, \ldots, u_k > 0$, and $\boldsymbol{y} \in \Sigma(\mathcal{A})$. By Taylor series expansion of each matrix exponential and equating coefficients, this is equivalent to

$$\boldsymbol{p}^T I_{y_0} Q^{j_1} I_{y_1} Q^{j_2} \cdots I_{y_{k-1}} Q^{j_k} I_{y_k} \mathbf{1} = \tilde{\boldsymbol{p}}^T \tilde{I}_{y_0} \tilde{Q}^{j_1} \tilde{I}_{y_1} \tilde{Q}^{j_2} \cdots \tilde{I}_{y_{k-1}} \tilde{Q}^{j_k} \tilde{I}_{y_k} \tilde{\mathbf{1}} \tag{19}$$

for all $k \geq 0$, $j_1, \ldots, j_k \geq 0$, and $\boldsymbol{y} \in \Sigma(\mathcal{A})$. Letting $j_i = 1$ for all i shows that

$$\boldsymbol{p}^T Q^{\boldsymbol{y}} \mathbf{1} = \tilde{\boldsymbol{p}}^T \tilde{Q}^{\boldsymbol{y}} \tilde{\mathbf{1}} \quad \text{for all } \alpha \in \Sigma(\mathcal{A}) \tag{20}$$

and the first direction is shown.

Now assume (20) holds. Using (11),

$$\begin{aligned}
\boldsymbol{p}^T I_{y_0} Q^{j_1} \cdots I_{y_{k-1}} Q^{j_k} I_{y_k} \mathbf{1} &= \sum \boldsymbol{p}^T I_{y_0} Q^{\boldsymbol{z}_1} \cdots I_{y_{k-1}} Q^{\boldsymbol{z}_k} I_{y_k} \mathbf{1} \\
&= \sum \tilde{\boldsymbol{p}}^T \tilde{I}_{y_0} \tilde{Q}^{\boldsymbol{z}_1} \cdots \tilde{I}_{y_{k-1}} \tilde{Q}^{\boldsymbol{z}_k} \tilde{I}_{y_k} \tilde{\mathbf{1}} \\
&= \tilde{\boldsymbol{p}}^T \tilde{I}_{y_0} \tilde{Q}^{j_1} \cdots \tilde{I}_{y_{k-1}} \tilde{Q}^{j_k} \tilde{I}_{y_k} \tilde{\mathbf{1}}
\end{aligned}$$

where $\sum$ represents a multiple sum over all sequences $|\boldsymbol{z}_i| = j_i$ for $1 \leq i \leq k$, and the second equality is a consequence of (20). Since (19) is equivalent to (18), this completes the proof. □

References

[1] Pierre Baldi, Yves Chauvin, Tim Hunkapiller, and Marcella A. McClure. Hidden Markov models of biological primary sequence information. *Proceedings of the National Academy of Sciences*, 91:1059–1063, 1994.

[2] David Blackwell and Lambert Koopmans. On the identifiability problem for functions of finite Markov chains. *The Annals of Mathematical Statistics*, 28:1011–1015, 1957.

[3] Joseph Felsenstein and Gary A. Churchill. A hidden Markov model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution*, 13(1):93–104, 1996.

[4] Donald R. Fredkin and John A. Rice. Identification of aggregated Markovian models: Application to the nicotinic acetylcholine receptor. In Lucien M. Le Cam and Richard A. Olshen,

editors, *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume I, pages 269–289. Wadsworth, Inc., 1985.

[5] Donald R. Fredkin and John A. Rice. On aggregated Markov processes. *Journal of Applied Probability*, 23:208–214, 1986.

[6] Edgar J. Gilbert. On the identifiability problem for functions of finite Markov chains. *The Annals of Mathematical Statistics*, 30:688–697, 1959.

[7] Hisashi Ito, Shun ichi Amari, and Kingo Kobayashi. Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE Transactions on Information Theory*, 38(2):324–333, March 1992.

[8] Paul Kienker. Equivalence of aggregated Markov models of ion-channel gating. *Proceedings of the Royal Society of London B*, 236:269–309, 1989.

[9] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.

[10] Tobias Rydén. On identifiability and order of continuous-time hidden Markov chains and Markov modulated Poisson processes. Lund University and Lund Institute of Technology, 1993.

[11] Walter Zucchini and Peter Guttorp. A hidden Markov model for space-time precipitation. *Water Resources Research*, 27(8):1917–1923, August 1991.

Bret Larget
Department of Mathematics and Computer Science
Duquesne University
440 College Hall
Pittsburgh, PA 15282-1707