A regularized spectral algorithm for Hidden Markov Models with applications in computer vision - Supplementary Material

Hà Quang Minh¹ Marco Cristani¹ Alessandro Perina² Vittorio Murino¹
¹ Istituto Italiano di Tecnologia (IIT), Genoa 16163, Italy
² Microsoft Research, WA, USA
{minh.ha quam,marco.cristani,vittorio.murino}@iit.it, alperina@microsoft.com

Abstract

The supplementary material contains the proofs of Lemmas 1, 2, 3 and 4 in the paper. The proofs for Proposition 1, Lemma 5, and Corollary 1 are given in the main paper.

We provide the proofs of Lemmas 1 and 2 for completeness. The proofs of Lemmas 3 and 4 are adapted from proof techniques in [1] and [2], respectively.

For clarity, we restate all the lemmas that we will prove here. For each symbol $x \in \{1, \ldots, n\}$, $A_x$ is defined by

$$A_x = T \text{diag}(O_{x,1}, \ldots, O_{x,m}) = TO_x \in \mathbb{R}^{m \times m},$$
with

$$O_x = \text{diag}(O_{x,1}, \ldots, O_{x,m}) \in \mathbb{R}^{m \times m}. \quad \text{(2)}$$

**Lemma 1.** The joint probability of a sequence $x_1, \ldots, x_t$, is given by

$$P(X_1 = x_1, \ldots, X_t = x_t) = 1^T_m A_x \ldots A_{x_1} \pi,$$
where $1_m = (1, \ldots, 1)^T$ is a column vector in $\mathbb{R}^m$.

**Proof of Lemma 1.** We first prove that

$$(A_x \ldots A_{x_1} \pi)_i = P(H_{t+1} = i, X_t = x_t, \ldots, X_1 = x_1). \quad \text{(4)}$$

The case $t = 1$: we have, using conditional independence

$$(A_x \pi)_i = \sum_{j=1}^m (A_x)_{ij} \pi_j = \sum_{j=1}^m P(H_2 = i | H_1 = j) P(X_1 = x_1 | H_1 = j) P(H_1 = j) = \sum_{j=1}^m P(H_2 = i, X_1 = x_1 | H_1 = j) P(H_1 = j) = P(H_2 = i, X_1 = x_1).$$

The case $t = 2$: we have, using conditional independence

$$(A_x A_{x_1} \pi)_i = \sum_{j=1}^m (A_x)_{ij} (A_{x_1} \pi)_j = \sum_{j=1}^m P(H_3 = i | H_2 = j) P(X_2 = x_2 | H_2 = j) P(H_2 = j) P(H_1 = x_1) = \sum_{j=1}^m P(H_3 = i, X_2 = x_2 | H_2 = j) P(H_2 = j, X_1 = x_1) = \sum_{j=1}^m P(H_3 = i, H_2 = j, X_2 = x_2, X_1 = x_1) = \sum_{j=1}^m P(H_3 = i, H_2 = j, X_2 = x_2, X_1 = x_1) = \sum_{j=1}^m P(H_3 = i, X_2 = x_2, X_1 = x_1).$$

In general, by induction, we get

$$(A_x \ldots A_{x_1} \pi)_i = P(H_{t+1} = i, X_t = x_t, \ldots, X_1 = x_1).$$

Then $1^T_m A_x \ldots A_{x_1} \pi = \sum_{i=1}^m (A_x \ldots A_{x_1} \pi)_i = P(X_t = x_t, \ldots, X_1 = x_1)$, as desired. \hfill \Box
Lemma 2. For each $1 \leq i \leq n$:

$$
\mathbb{P}(X_{t+1} = i, X_t = x_t, \ldots, X_1 = x_1) = (OA_{x_1} \ldots A_{x_t} \pi_t)_{ij}.
$$

(5)

$$
\mathbb{P}(X_{t+1} = i|X_t = x_t, \ldots, X_1 = x_1) = \frac{(OA_{x_1} \ldots A_{x_t} \pi_t)_{ij}}{\mathbb{P}(X_{t+1})}.
$$

(6)

Proof of Lemma 2. Using conditional independence and the proof of Lemma 1, we get

$$
(\sum_{j=1}^m O_{ij} (A_{x_1} \ldots A_{x_t} \pi_t)_{ij}) = \sum_{j=1}^m O_{ij} (A_{x_1} \ldots A_{x_t} \pi_t)_{ij}
$$

$$
= \sum_{j=1}^m \mathbb{P}(X_{t+1} = i|H_{t+1} = j) \mathbb{P}(H_{t+1} = j, X_t = x_t, \ldots, X_1 = x_1)
$$

$$
= \sum_{j=1}^m \mathbb{P}(X_{t+1} = i|H_{t+1} = j, X_t = x_t, \ldots, X_1 = x_1) \mathbb{P}(H_{t+1} = j, X_t = x_t, \ldots, X_1 = x_1)
$$

$$
= \mathbb{P}(X_{t+1} = i|X_t = x_t, \ldots, X_1 = x_1)
$$

as desired. □

Lemma 3. Under Assumptions 1, consider the following quantities:

$$
\begin{align*}
    b_{\infty} &= (U^T P_{2,1} P_{2,1}^T U)^{-1}(U^T P_{2,1} P_1), \\
    B_x &= (U^T P_{3,1} P_{2,1}^T U)(U^T P_{2,1} P_{2,1}^T U)^{-1}, \\
    b_1 &= U^T P_1 = (U^T D\pi)^T
\end{align*}
$$

Then

$$
\begin{align*}
    b_{\infty}^T &= 1_m^T (U^T O)^{-1}, \\
    B_x &= (U^T O)A_x(U^T O)^{-1},
\end{align*}
$$

and consequently

$$
\begin{align*}
    b_{\infty}^T B_{x1} b_1 &= 1_m^T A_{x1} \pi_1.
\end{align*}
$$

(10)

Before proving Lemma 3, recall that by Assumptions 1, $O$ is of full column rank, and both $T$ and $\text{diag}(\pi)$ are invertible, so that

$$
\text{rank}(P_{2,1}) = \text{rank}(O\text{diag}(\pi)O^T) = \text{rank}(T\text{diag}(\pi)O^T) = \text{rank}(O^T) = m.
$$

Similarly, since $U^T O$ is invertible,

$$
\text{rank}(U^T P_{2,1} P_{2,1}^T U) = \text{rank}(U^T P_{2,1}) = \text{rank}(O^T) = m.
$$

This shows that the $m \times m$ matrix $U^T P_{2,1} P_{2,1}^T U$ is invertible. Conversely, if $U^T P_{2,1} P_{2,1}^T U$ is of full rank $m$, then $U^T O$ is also of full rank $m$. Thus the quantities $B_x, b_{\infty}$ are well-defined.

Proof of Lemma 3. First note that

$$
P_1 = 1_n^T P_{2,1} = 1_n^T O \text{diag}(\pi) O^T = 1_m^T O \text{diag}(\pi) O^T,
$$

since $1_n^T O = 1_n^T$. We then have

$$
\begin{align*}
P_1^T P_{2,1} U &= 1_m^T T \text{diag}(\pi) O^T P_{2,1}^T U \\
&= 1_m^T (U^T O)^{-1} (U^T O) T \text{diag}(\pi) O^T P_{2,1}^T U \\
&= 1_m^T (U^T O)^{-1} (U^T P_{2,1} P_{2,1}^T U).
\end{align*}
$$

Thus

$$
\begin{align*}
b_{\infty}^T &= (P_1^T P_{2,1}^T U)(U^T P_{2,1} P_{2,1}^T U)^{-1} \\
&= 1_m^T (U^T O)^{-1} (U^T P_{2,1} P_{2,1}^T U)(U^T P_{2,1} P_{2,1}^T U)^{-1} \\
&= 1_m^T (U^T O)^{-1}.
\end{align*}
$$
Next, we have
\[ P_{3,x,1} = O_A x T \text{diag}(\mathbf{P}) O_T \]
\[ = O_A x (U^T O)^{-1} (U^T O) T \text{diag}(\mathbf{P}) O_T \]
\[ = O_A x (U^T O)^{-1}U^T P_{2,1}, \]
so that
\[ U^T P_{3,x,1} P_{2,1}^T U = (U^T O)A_x(U^T O)^{-1}U^T P_{2,1} P_{2,1}^T U. \]

Therefore
\[ B_x = (U^T P_{3,x,1} P_{2,1}^T U)(U^T P_{2,1} P_{2,1}^T U)^{-1} \]
\[ = (U^T O)A_x(U^T O)^{-1}. \]
It follows immediately that
\[ b_\infty^T B_{x,1} b_1 = 1_m^T A_{x,1} \mathbf{P}. \]

This completes the proof.

\[ \square \]

**Lemma 4.** Consider the following quantities:

\[ b_\infty = (U^T P_{2,1} P_{2,1}^T U)^{-1}(U^T P_{2,1} P_{2,1}^T U) \in \mathbb{R}^k, \]

\[ (11) \]
\[ B_x = (U^T P_{3,x,1} P_{2,1}^T U)(U^T P_{2,1} P_{2,1}^T U)^{-1} \in \mathbb{R}^{k \times k}, \]

\[ (12) \]
Then
\[ b_\infty^T = 1_m^T R(U^T OR)^{-1}, \]
\[ (13) \]
\[ B_x = (U^T OR)S_x R(U^T OR)^{-1}, \]
and consequently
\[ b_\infty^T B_{x,1} b_1 = 1_m^T A_{x,1} \mathbf{P}. \]

Before proving Lemma 4, recall that we assume that there is a matrix \( U \in \mathbb{R}^{n \times k} \) such that
\[ \text{rank}(U^T OR) = \text{rank}(S \text{diag}(\mathbf{P}) O_T) = k. \]

Since \( U^T OR \) is a \( k \times k \) matrix and \( P_{2,1} = ORS \text{diag}(\mathbf{P}) O_T \), this implies that
\[ \text{rank}(U^T P_{2,1} P_{2,1}^T U) = \text{rank}(U^T P_{2,1}) = \text{rank}(S \text{diag}(\mathbf{P}) O_T) = k. \]

Thus the \( k \times k \) matrix \( (U^T P_{2,1} P_{2,1}^T U) \) is invertible, showing that \( B_x \) and \( b_\infty \) are well-defined. Conversely, if \( \text{rank}(U^T P_{2,1} P_{2,1}^T U) = k \), then \( U^T OR \) is of full rank \( k \), hence invertible. Furthermore, \( \text{rank}(U^T OR) = k \Rightarrow \text{rank}(OR) \geq \text{rank}(U^T OR) = k. \)

On the other hand, \( OR \in \mathbb{R}^{n \times k} \), with \( k \leq n \), so that \( \text{rank}(OR) \leq k. \) Thus we have
\[ \text{rank}(OR) = k. \]

Since \( OR \) is now a matrix of full column rank, we have
\[ \text{rank}(P_{2,1}) = \text{rank}(ORS \text{diag}(\mathbf{P}) O_T) = \text{rank}(S \text{diag}(\mathbf{P}) O_T) = k. \]

This reasoning clearly applies both for \( k \leq m \leq n \) and \( k \leq n \leq m \), which is what we noted at the beginning of Section 3.5.

**Proof of Lemma 4.** We have
\[ P_{2,1}^T U = 1_m^T T \text{diag}(\mathbf{P}) O_T P_{2,1}^T U \]
\[ = 1_m^T RS \text{diag}(\mathbf{P}) O_T P_{2,1}^T U \]
\[ = 1_m^T R(U^T OR)^{-1}(U^T OR) S \text{diag}(\mathbf{P}) O_T P_{2,1}^T U \]
Thus

\[
\begin{align*}
1^T_m R (U^T OR)^{-1} (U^T) O^T \text{diag}(\vec{\pi}) O T \bar{P}_k^T U & \\
= 1^T_m R (U^T OR)^{-1} (U^T P_{2,1} \bar{P}_k^T U).
\end{align*}
\]

Next, we have

\[
\begin{align*}
P_{3,x,1} &= OA_x T \text{diag}(\vec{\pi}) \bar{O} T = OA_x R S \text{diag}(\vec{\pi}) O T \\
&= OA_x R (U^T OR)^{-1} (U^T OR) S \text{diag}(\vec{\pi}) O T \\
&= OA_x R (U^T OR)^{-1} U^T = ORS O_x R (U^T OR)^{-1} U^T P_{2,1},
\end{align*}
\]

so that

\[
U^T P_{3,x,1} P_{2,1}^T U = (U^T OR) S O_x R (U^T OR)^{-1} U^T P_{2,1} P_{2,1}^T U
\]

Therefore

\[
B_x = (U^T P_{3,x,1} P_{2,1}^T U) (U^T P_{2,1} P_{2,1}^T U)^{-1}
\]

\[
= (U^T OR) S O_x R (U^T OR)^{-1}. \quad \text{For each } x \text{ we have}
\]

\[
b^T_x B_x b_1 = [1^T_m R (U^T OR)^{-1}] [(U^T OR) S O_x R (U^T OR)^{-1}] [U^T OR \vec{\pi} k]
\]

\[
= 1^T_m R S O_x R \vec{\pi} k = 1^T_m T O_x \vec{\pi} = 1^T_m A_x \vec{\pi}. \quad \text{This completes the proof.}
\]

References
