Correction to
“A Generalization of the Blahut-Arimoto Algorithm to Finite-State Channels” by Vontobel et al.

Pascal O. Vontobel, Senior Member, IEEE, Aleksandar Kavčić, Senior Member, IEEE,
Dieter M. Arnold, Member, IEEE, Hans-Andrea Loeliger, Fellow, IEEE

Abstract—The classical Blahut–Arimoto algorithm (BAA) optimizes a discrete memoryless source (DMS) at the input of a discrete memoryless channel (DMC) in order to maximize the mutual information between the channel input and output. The paper [Vontobel, Kavčić, Arnold, Loeliger, IEEE Trans. Inf. Theory, May 2008] presented a generalization of the BAA, called the generalized BAA (GBAA), for optimizing finite-state machine sources (FSMSs) at the input of finite-state machine channels (FSMCs) in order to optimize the mutual information rate between the channel input and output. We became aware of the fact that a step in the proof of one of the results is only valid for certain FSMS/FSMC pairs. In this note we present a correction to that result (and its proof) so that it holds for any FSMS/FSMC pair under consideration. We also state sufficient conditions on FSMS/FSMC pairs for which this correction is not required.

I. INTRODUCTION

The paper [1] introduced the generalized Blahut–Arimoto algorithm (GBAA) for optimizing finite-state machine sources (FSMSs) at the input of finite-state machine channels (FSMCs) in order to optimize the mutual information rate between the channel input and output. A key ingredient for presenting the results in [1] were the so-called $T$-values that were defined in Definition 41 of [1]. On the one hand, these $T$-values were used to express various entropy and information rates, on the other hand, these $T$-values were used to express derivatives of various entropy and information rates w.r.t. FSMS parameters. With this, the $T$-values play also a central role in the formulation of the GBAA in Algorithm 45 of [1]. As we found out, a step in the proof of Lemma 60 of [1] is, unfortunately, only valid if the FSMS/FSMC pair satisfies a certain condition, i.e., the result was not as generally applicable as claimed in Lemma 60 of [1].

The present note contains the following results.

- **We discuss sufficient conditions for which the $T$-value correction terms are zero.** In particular, we highlight two important classes of FSMS/FSMC pairs for which these correction terms are zero:
  1) FSMS/FSMC pairs where the FSMS state at a given time index uniquely determines the FSMC state at the same time index;
  2) FSMCs where the FSMC state sequence is independent of the sequence at its input.

Given this, the results and derivations in [2], [3] are, to the best of our knowledge, correct because the FSMS/FSMC pairs that were considered in [2], [3] fall into the first category. Moreover, the results presented in Examples 71–73 of [1] are also correct because the FSMS/FSMC pairs in Examples 71 and 72 fall into the first category and the FSMS/FSMC pairs in Example 73 fall into the second category.

- **When computing the entropy and information rates with the help of the $T$-values, adding the $T$-value correction terms is not required** because, to the best of our knowledge, the proofs in [1] were correct with the original definition of the $T$-values. However, interestingly, the entropy and information rate expressions are also correct with the modified $T$-values; this stems from the fact that the relevant weighted average of the $T$-value correction terms is always zero for the entropy and information rates.

For the rest of the paper, we assume that the reader is familiar with the basics of the nomenclature and the notation of [1]. Moreover, if not mentioned otherwise, then all definition, lemma, theorem, example, and algorithm numbers refer to the corresponding items in [1]. However, if not stated otherwise, equation numbers refer to equation numbers in this note.

II. $T$-VALUE CORRECTION TERMS

In this section we explain which step in the proof of Lemma 60 requires the FSMS/FSMC pair to satisfy a certain condition for it to be true. We then present the $T$-value correction terms that are required so that the statement and the proof of Lemma 60 hold for any FSMS/FSMC pair under consideration. Finally, we discuss efficient ways of approximating the $T$-value correction terms.
\[ J = \frac{1}{2N} \sum_{(i,j) \in B} \sum_{b_{(i,j)} \in B} Q_{ij}^a \sum_{\ell \in I_N} \sum_{b_{(i,j)}} \frac{Q(b)}{Q_{ij}} \sum_{b''} W(b'', y|b) \log \left( \frac{V(b''|y)}{V(b''|b, y)} \right) \]

\[ J = \frac{1}{2N} \sum_{(i,j) \in B} \sum_{b_{(i,j)} \in B} Q_{ij}^a \sum_{\ell \in I_N} \sum_{b_{(i,j)} \in B} \frac{Q(b)}{Q_{ij}} \sum_{b''} W(b'', y|b) \log \left( \frac{V(b''|y)}{V(b''|b, y)} \right) \]

**A. The Mistake**

Consider the following two expressions for \( J \):

- Eq. (1) at the top of this page (which reproduces the expression in the middle of Page 1915 of [1]);
- Eq. (2) in the middle of this page (which reproduces the expression at the top of Page 1916 of [1]).

Paper [1] claimed that (2) follows from (1), in particular that Lines 1, 2, 3, 4, 5, 6 in (2) follow from, respectively, Lines 1, 2, 3, 4, 5, 6 in (1). Unfortunately, Line 2 in (2) in general does not follow from Line 2 of (1). Similarly, Line 5 in (2) in general does not follow from Line 5 of (1). However, Lines 1, 3, 4, and 6 in (2) do follow from the corresponding lines in (1).

The problem is that the transition from Lines 2 and 5 in (1) to Lines 2 and 5 in (2) requires several simplifications, one of them replacing

\[ W(b'^N_{\ell+1}, y^N_{\ell+1}|b) \text{ by } W(b'^N_{\ell+1}, y^N_{\ell+1}|b_{\ell+1}). \]

For general FSMS/FSMC pairs, this simplification is not allowed because of the dependence of \( b'^N_{\ell+1} \) and \( y^N_{\ell+1} \) on \( b'_{\ell+N} \), even when conditioned on \( b_{\ell+1}^N \).

The proof of Lemma 60 then proceeded by showing that Lines 2 and 5 of (2) cancel each other and that Lines 3 and 6 of (2) cancel each other, so the final result was effectively
given by Lines 1 and 4 of (2). However, once the Lines 2 and 5 in (2) are corrected, they in general do not cancel anymore and this non-cancellation will lead to the above-mentioned $T$-value correction terms.

#### B. Correction of the Mistake

We start the presentation of the correction of this mistake by defining $J_2$ and $J_5$ as shown in (3) and (4), respectively, at the top of this page. These expressions represent, respectively, Lines 2 and 5 of (1). (Recall that $b_{\ell t} = (i,j)$ means that $s_{\ell-1} = i$ and $s_\ell = j$.)

In a first step, we reformulate the expression for $J_5$ as shown in (5) near the top of this page, where step (a) follows from $\sum_{(i,j) \in B} Q_{ij}^b f(i) = \sum_{i \in S} \sum_{j \in B} Q_{ij}^b f(i) = \sum_{i \in S} h_i^b f(i)$ (here, $f(i)$ is an expression that depends only on $i$, but not on $j$), where step (b) follows from replacing $i$ by $j$, and where step (c) follows from a similar argument as in step (a). (Note that these transformation steps are similar to the transformation steps that are applied to the expression near the top of the second column of Page 1914 of [1].)

Let $\tilde{J} \triangleq J_2 + J_5$. In the following, we want to simplify $\tilde{J}$, which represents the contribution of $J_2$ and $J_5$ to $J$. First, we observe that in (3) we can change the summation range from $\ell \in I_N$ to $\ell \in I_N'$ because the contribution of the $\ell = N$ term is zero. Afterwards, combining (3) and (5), we obtain

\[
J_2 \triangleq \frac{1}{2N} \sum_{(i,j) \in B} \sum_{\ell \in I_N} Q_{ij}^b \sum_{b' \in B} \frac{Q(b)}{Q_{ij}} \sum_{y} W(b', y|b) \log \left( V(b_{\ell+1}^N s_{\ell+1}^y, y_{\ell+1}^N) \right) \tag{3}
\]

\[
J_5 \triangleq \frac{1}{2N} \sum_{b} \sum_{(i,j) \in B} \sum_{\ell \in I_N} Q_{ij}^b \sum_{b' \in B} \frac{Q(b)}{\mu_i} \sum_{y} W(b', y|b) \log \left( V(b_{\ell+1}^N s_{\ell+1}^y, y_{\ell+1}^N) \right) \tag{4}
\]

where we have defined\(^1\)

\[
\hat{G}_{ij}(\ell, s_{\ell}^y) \triangleq \sum_{b_{\ell-1}^N} \frac{Q(b_{\ell-1}^N | b_{\ell}^b)}{Q_{ij}} \cdot W(s_{\ell}^y | b_{\ell-1}^N) - \sum_{b_{\ell-1}^N} \frac{Q(b_{\ell-1}^N | b_{\ell}^b)}{\mu_j} \cdot W(s_{\ell}^y | b_{\ell-1}^N) \tag{7}
\]

and

\[
\hat{F}(\ell, s_{\ell}^y) \triangleq \sum_{b_{\ell+1}^N} Q(b_{\ell+1}^N | s_{\ell}) \cdot \left( \sum_{b_{\ell+1}^N} W(b_{\ell+1}^N | s_{\ell+1}^y, y_{\ell+1}^N | s_{\ell}^y, b_{\ell+1}^N) \cdot \log \left( V(b_{\ell+1}^N | s_{\ell}^y, y_{\ell+1}^N) \right) \right) \tag{8}
\]

Recall that $s_{\ell}^y = (s_{\ell}, s_{\ell}^y)$, i.e., $s_{\ell}^y$ specifies the FSM state $s_{\ell}$ and the FSMC state $s_{\ell}^y$ at time index $\ell$.

One observes that the function $\hat{F}(\ell, s_{\ell}^y)$ is closely related to $\chi_{s_{\ell}}(\ell)$ that is defined in Eq. (54) of Page 1916 of [1]. However, unlike $\hat{F}(\ell, s_{\ell}^y)$, the function $\chi_{s_{\ell}}(\ell)$ does not depend on $s_{\ell}^y$.

Moreover, note that $\hat{G}_{ij}(\ell, s_{\ell}^y)$ can also be written as

\[
\hat{G}_{ij}(\ell, s_{\ell}^y) = \sum_{b_{\ell-1}^N \neq b_{\ell}^b} \frac{Q(b_{\ell-1}^N | b_{\ell}^b)}{Q_{ij}} \cdot W(s_{\ell}^y | b_{\ell-1}^N) - \sum_{b_{\ell-1}^N = b_{\ell}^b} \frac{Q(b_{\ell-1}^N | b_{\ell}^b)}{\mu_j} \cdot W(s_{\ell}^y | b_{\ell-1}^N) \tag{9}
\]

\(^1\)In the same way that the proof of Lemma 60 omitted the superscript $(N)$ to $J$ for reasons of simplicity, we omit here the the superscript $(N)$ to $\hat{G}_{ij}(\ell, s_{\ell}^y)$ and $\hat{F}(\ell, s_{\ell}^y)$. 

\[
\tilde{J} = \frac{1}{2N} \sum_{(i,j) \in B} \sum_{\ell \in I_N \setminus s_{\ell}^y} \hat{G}_{ij}(\ell, s_{\ell}^y) \cdot \hat{F}(\ell, s_{\ell}^y), \tag{6}
\]
C. Correction Terms

With the development in the previous subsection, it is straightforward to define the correction terms that need to be added to the expressions of the $T$-values in Definition 41. Namely, let

$$
\overrightarrow{T}_{ij} \triangleq \lim_{N \to \infty} \overrightarrow{T}^{(N)}_{ij},
$$

(10)

$$
\overrightarrow{T}^{(N)}_{ij} \triangleq \frac{1}{2N} \sum_{\ell \in I_{ij}} \overrightarrow{T}^{(N)}_{ij}(\ell),
$$

(11)

$$
\overrightarrow{T}^{(N)}_{ij}(\ell) \triangleq \sum_{s_i'} \sum_{s_i''} \overrightarrow{G}_{ij}(\ell, s_i'); \overrightarrow{F}(\ell, s_i''),
$$

(12)

where $\overrightarrow{G}_{ij}(\ell, s_i'')$ and $\overrightarrow{F}(\ell, s_i'')$ were defined in (7) and (8), respectively.

- $\overrightarrow{T}_{ij}$ has to be added to the expression for $\overrightarrow{T}_{ij}$,
- $\overrightarrow{T}^{(N)}_{ij}$ has to be added to the expression for $\overrightarrow{T}_{ij}$,
- $\overrightarrow{T}^{(N)}_{ij}(\ell)$ has to be added to the expression for $\overrightarrow{T}_{ij}(\ell)$.

(Note that we have formulated the correction terms so that there is only a correction to $\overrightarrow{T}_{ij}$, etc., but no correction to $\overrightarrow{T}_{ij}$, etc. Things could also have been formulated such that there are corrections to both $\overrightarrow{T}_{ij}$, etc., and $\overrightarrow{T}_{ij}$, etc.)

With these $T$-value correction terms, the expressions of the derivatives of the entropy and information rate in Lemma 60 and Theorem 66, along with the GBAA expressions in Algorithm 45, are correct for general FSMS/FSMC pairs.

D. Entropy and Information Rates

Lemma 59 and Theorem 65 (and their proofs) about the values of certain entropy and information rates are, to the best of our knowledge, correct with the $T$-values as originally defined in Definition 41. However, interestingly enough, the expressions in Lemma 59 and Theorem 65 are still correct when the $T$-values in Definition 41 are modified with the above correction terms. This observation stems from the fact that, although the $T$-value correction terms might be nonzero, the weighted average of the $T$-value correction terms that appears in Lemma 59 and Theorem 65 is always zero. Namely,

$$
\sum_{(i,j) \in B} Q_{ij} \overrightarrow{T}^{(N)}_{ij}
= \sum_{(i,j) \in B} Q_{ij} \frac{1}{2N} \sum_{\ell \in I_{ij}} \sum_{s_i''} \overrightarrow{G}_{ij}(\ell, s_i''); \overrightarrow{F}(\ell, s_i'')
= \frac{1}{2N} \sum_{\ell \in I_{ij}} \sum_{(i,j) \in B} Q_{ij} \overrightarrow{G}_{ij}(\ell, s_i''); \overrightarrow{F}(\ell, s_i'')
= 0,
$$

(a)

where step (a) used $\sum_{(i,j) \in B} Q_{ij} \overrightarrow{G}_{ij}(\ell, s_i'') = 0$. This latter expression follows from

$$
\sum_{(i,j) \in B} Q_{ij} \overrightarrow{G}_{ij}(\ell, s_i'') =
= \sum_{(i,j) \in B} \sum_{b_{ij} = (i,j)} Q(b_{ij}^{\ell+1}) \cdot W(s_{i'}\mid b_{ij}^{\ell+1})
- \sum_{(i,j) \in B} Q_{ij} b_{ij}^{\ell+1} \cdot W(s_{i'}\mid b_{ij}^{\ell+1})
= \sum_{b_{ij}^{\ell+1} \mid s_{i'} \mid j} Q(b_{ij}^{\ell+1}) \cdot W(s_{i'}\mid b_{ij}^{\ell+1})
= 0,
$$

where in step (b) we have used (7) and where in step (c) we have used $\sum_{i,j} Q_{ij} f(j) = \sum_{j \in S} \sum_{i,j} Q_{ij} f(j) = \sum_{j \in S} \sum_{ij} f(j)$, where $f(j)$ is an expression that depends only on $j$, but not on $i$.

E. Cases where the $T$-Value Correction Terms are Zero

For some FSMS/FSMC pairs, the $T$-value correction terms turn out to be zero. A sufficient condition for this to happen is that there is some function $\overrightarrow{g} : S \times S' \to \mathbb{R}$ such that

$$
W(s_{i'}\mid b_{ij}^{\ell+1}) = \overrightarrow{g}(s_{i'}, s_{i'})
$$

(13)

for all $s_{i'}$ and $b_{ij}^{\ell+1}$. This follows from the observation that

$$
\overrightarrow{G}_{ij}(\ell, s_{i'}') \triangleq \sum_{b_{ij}^{\ell+1}} Q(b_{ij}^{\ell+1} \mid b_{ij}) \cdot \overrightarrow{g}(s_{i'}, s_{i'})
- \sum_{b_{ij}^{\ell+1} \mid s_{i'} \mid j} Q(b_{ij}^{\ell+1} \mid b_{ij}) \cdot \overrightarrow{g}(s_{i'}, s_{i'})
= \left( \sum_{b_{ij}^{\ell+1}} Q(b_{ij}^{\ell+1} \mid b_{ij}) \right) \cdot \overrightarrow{g}(s_{i'}, s_{i'})
- \sum_{b_{ij}^{\ell+1} \mid s_{i'} \mid j} Q(b_{ij}^{\ell+1} \mid b_{ij}) \cdot \overrightarrow{g}(s_{i'}, s_{i'})
= 0,
$$

where at step (a) we have used (9), $s_{i'}' = (s_{i'}, s_{i'})$, $W(s_{i'}'\mid b_{ij}^{\ell+1}) = W(s_{i'}'\mid b_{ij}^{\ell+1})$, and (13). Combining this with (10)–(12) yields the anticipated result that $\overrightarrow{T}^{(N)}_{ij}(\ell) = 0$, for all $(i,j) \in B$ and $\ell \in I_{ij}$, along with $\overrightarrow{T}_{ij} = 0$ and $\overrightarrow{T}^{(N)}_{ij} = 0$ for all $(i,j) \in B$.

Let us now discuss two important classes of FSMS/FSMC pairs where the sufficient condition in (13) is satisfied.
1) FSMS/FSMC pairs where the FSMS state at a given time index uniquely determines the FSMC state at the same time index, i.e., FSMS/FSMC pairs for which there is some function \( g' : S \to S' \) such that \( s'_\ell = g'(s_\ell) \) for every \( s_\ell \in S \). With this, the function \( \overline{g} \) in (13) is given by \( \overline{g}(s_\ell, s'_\ell) \triangleq [g'(s_\ell) = s'_\ell] \), where we have used Iverson’s notation, i.e., \( |S| = 1 \) if the statement \( S \) is true, and \( |S| = 0 \) otherwise.

This class of FSMS/FSMC pairs includes, for example, the case where the FSMC is controllable (see Remark 21 and Definition 22) and the FSMS state space is a superset of the FSMC state space. (Here we neglect the fact that a bounded number of time indices might be needed at the beginning so that the FSMC state is given by the FSMS state.)

2) FSMCs where the FSMS state sequence is independent of the sequence at its input. With this, the function \( \overline{g} \) in (13) is given by \( \overline{g}(s_\ell, s'_\ell) \triangleq \Pr(S'_\ell = s'_\ell) \).

This class of FSMS/FSMC pairs includes, for example, the case where the FSMC is a Gilbert–Elliott channel (see Example 20). (Note that no restrictions need to be imposed on the choice of the FSMS.)

Given this, the results and derivations in [2, 3] are, to the best of our knowledge, correct because the FSMS/FSMC pairs that were considered in [2, 3] fall into the first category. Moreover, the results presented in Examples 71–73 of [1] are also correct because the FSMS/FSMC pairs in Examples 71 and 72 fall into the first category and the FSMS/FSMC pairs in Example 73 fall into the second category.

Of course, there are also other FSMS/FSMC pairs where \( \overline{T}_{ij} = 0 \) for all \((i, j) \in B\). However, in these cases usually cancellations happen due to more subtle symmetries.

**F. Comments on Lines 3 and 6 in Eq. (1)**

As mentioned at the end of Section II-A, Lines 3 and 6 of (1) cancel. The derivation of this fact was, to the best of our knowledge, correct in [1]. However, we can re-derive it by suitably modifying the formalism that was used in the context of Lines 2 and 5.

Towards this end, let \( J_3 \) and \( J_6 \) represent Lines 3 and 6 of (1), respectively, and let \( J = J_3 + J_6 \). The function analogous to \( \overline{G}_{ij}(\ell, s'_\ell) \) is

\[
\overline{\tilde{G}}_{ij}(\ell, s'_\ell) \triangleq \sum_{b_{\ell+1}^{N-1}} Q(b_{\ell+1}^{N-1} | b_\ell) \cdot W(s'_{\ell-1} | b_\ell^N) - \sum_{s_{\ell-1} = i} Q(b_\ell^N | s_{\ell-1}) \cdot W(s'_{\ell-1} | b_\ell^N).
\]

However, given the FSMC models under consideration, we have \( W(s'_{\ell-1} | b_\ell^N) = W(s'_{\ell-1} | s_{\ell-1}) \). Therefore,

\[
W(s''_{\ell-1} | b_\ell^N) = W(s'_{\ell-1} | s_{\ell-1}), \text{ and so}
\]

\[
\overline{\tilde{G}}_{ij}(\ell, s'_\ell) \triangleq \left( \sum_{b_{\ell+1}^{N-1}} Q(b_{\ell+1}^{N-1} | b_\ell) \cdot W(s'_{\ell-1} | b_\ell^N) - \sum_{s_{\ell-1} = i} Q(b_\ell^N | s_{\ell-1}) \cdot W(s'_{\ell-1} | b_\ell^N) \right) \cdot W(s''_{\ell-1} | s_{\ell-1})
\]

\[
= 0.
\]

An expressions analogous to (6) then yields \( \overline{J} = 0 \), i.e., the promised result.

**G. Approximating the Expression in Eq. (6)**

In principle, the expression for \( \overline{J} \) in (6) could be approximated (to any desired accuracy) with the help of simulation-based computations. However, the complexity of such computations would scale as \( O(N^2) \). Towards obtaining an expression whose computational complexity scales as \( O(NM) \), where \( M \) is some positive integer that depends on the memory length of the FSMS/FSMC pair but not on \( N \), we need to better understand the expression for \( \overline{J} \).

In particular, note that \( \overline{J} \) is of order \( O(1) \), a fact which is a priori not clear because \( \overline{G}_{ij}(\ell, s'_\ell) \) is of order \( O(1) \) and \( \overline{F}(\ell, s'_\ell) \) is of order \( O(N) \). However, a significant amount of (near) cancellations happen in (6), as we will see in the following. Part of the reason for these (near) cancellations are the fact that

\[
\sum_{s''_\ell} \overline{G}_{ij}(\ell, s''_\ell) = 0
\]

and the fact that

\[
\sum_{s''_\ell} \overline{G}_{ij}(\ell, s''_\ell) \cdot \overline{F}(\ell, s''_\ell)
\]

\[
= \left( \sum_{s''_\ell} \overline{G}_{ij}(\ell, s''_\ell) \right) \cdot \left( \sum_{s''_\ell} \overline{F}(\ell, s''_\ell) \right)
\]

\[
= 0
\]

if \( \overline{F}(\ell, s''_\ell) \) does not depend on \( s''_\ell \).

In order to reformulate (6), it turns out to be useful to define \( \overline{F}(\ell, m, s''_\ell) \) as shown in (7) at the top of the next page. Here, \( \ell \in I_N \), the integer \( m \) is such that \( 1 \leq m \leq N - \ell \), and \( s''_\ell \in S'' \). With this,

\[
\overline{F}(\ell, s''_\ell) = \sum_{m=1}^{N-\ell} \overline{F}(\ell, m, s'_{\ell})
\]

and so

\[
\overline{J} = \frac{1}{2N} \sum_{(i, j) \in B} Q_{ij} \sum_{\ell \in I_N} \sum_{m=1}^{N-\ell} \overline{G}_{ij}(\ell, s''_\ell) \cdot \overline{F}(\ell, m, s''_\ell).
\]
\[
\bar{F}(\ell, m, s''_\ell) \triangleq \sum_{b_{\ell+1}^N} Q(b_{\ell+1}^N|s_{\ell}) \sum_{b_{\ell+1}^N} \sum_{y_{\ell+1}^N} W(b_{\ell+1}^N, y_{\ell+1}^N|s_{\ell}', b_{\ell+1}^N) \log \left( V(b_{\ell+m}^N|b_{\ell+m+1}^N, s''_{\ell}, y_{\ell+1}^N) \right)
\]

(7)

\[
\bar{F}^{[M]}(\ell, s''_\ell) = \sum_{b_{\ell+1}^N} Q(b_{\ell+1}^N|s_{\ell}) \sum_{b_{\ell+1}^N} \sum_{y_{\ell+1}^N} W(b_{\ell+1}^N, y_{\ell+1}^N|s''_{\ell}, b_{\ell+1}^N) \log \left( V(b_{\ell+1}^{M+1}|b_{\ell+M+1}^N, s''_{\ell}, y_{\ell+1}^N) \right)
\]

(8)

Note that the term \( V(b_{\ell+m}^N|b_{\ell+m+1}^N, s''_{\ell}, y_{\ell+1}^N) \), which appears in \( \bar{F}(\ell, m, s''_{\ell}) \), can be expressed in terms of forward-backward algorithm computations. With this, and because the time index difference between \( s''_{\ell} \) and \( b_{\ell+m} \) grows linearly with \( m \), we see, for the FSMS/FSMC pairs under consideration, that the dependency of the term \( V(b_{\ell+m}^N|b_{\ell+m+1}^N, s''_{\ell}, y_{\ell+1}^N) \) on \( s''_{\ell} \) vanishes for increasing \( m \). In fact, for the considered FSMS/FSMC pairs the expression \( \sum s''_{\ell} \bar{G}_{ij}(\ell, s''_{\ell}) \cdot \bar{F}(\ell, m, s''_{\ell}) \) decays exponentially for \( m \to \infty \).

Therefore, choosing a suitably large integer \( M \) (that depends on the memory length of the FSMS/FSMC pair), we can approximate \( \bar{F} \) to any desired accuracy for \( N \) to any desired accuracy for \( N \) to infinity for \( \bar{F}(\ell, m, s''_{\ell}) \approx \sum_{s''_{\ell}} \bar{G}_{ij}(\ell, s''_{\ell}) \cdot \bar{F}(\ell, m, s''_{\ell}) \), where

\[
\bar{F}(\ell, m, s''_{\ell}) \triangleq \sum_{m=1}^{\min(M, N-\ell)} \bar{F}(\ell, m, s''_{\ell}).
\]

For the reader’s convenience, Eq. (8) at top of this page shows the expression for \( \bar{F}^{[M]}(\ell, s''_{\ell}) \) after replacing \( \bar{F}(\ell, m, s''_{\ell}) \) by the expression in (7). (Note that here and in the following we use \( \ell + M \) instead of the correct, but slightly more cumbersome, \( \min(\ell + M, N) \).)

H. Simulation-Based Approximation of (14)

There are various ways to approximate the expression in (14) to any desired accuracy for \( N \to \infty \) with the help of typical sequences \( \hat{b}'' \) and \( \hat{y} \) for \( b'' \) and \( y \), respectively. For example, one can use the following approximations.

Namely, note that \( \bar{G}_{ij}(\ell, s''_{\ell}) \) is essentially independent of \( \ell \) (except for \( \ell \) close to \( -N + 1 \) and \( +N \)), and so, in the limit \( N \to \infty \), we can approximate the expression \( \bar{G}_{ij}(\ell, s''_{\ell}) \) by the limit \( \bar{G}_{ij}(s''_{\ell}) \triangleq \lim_{N \to \infty} \frac{1}{2N} \sum_{\ell \in T_N} \bar{G}_{ij}(\ell, s''_{\ell}) \). In turn, the value of \( \bar{G}_{ij}(s''_{\ell}) \) can be approximated by

\[
\frac{1}{2NQ_{ij}} \sum_{\ell \in T_N} \left[ \hat{s}_{\ell}'' = s''_{ij} \right] - \frac{1}{2N\mu_j} \sum_{\ell \in T_N} \left[ \hat{s}_{\ell}'' = s'' \right],
\]

or, equivalently,

\[
\frac{1}{2NQ_{ij}} \sum_{\ell \in T_N} \left[ \hat{s}_{\ell-1} = i, \hat{s}_{\ell} = j \right] - \frac{1}{2N\mu_j} \sum_{\ell \in T_N} \left[ \hat{s}_{\ell} = j \right].
\]

(Here we have used Iverson’s notation.) Better approximations might be obtained by replacing \( Q_{ij} \) and \( \mu_j \) in the above expressions by the corresponding empirical averages based on the typical sequence \( \hat{b}'' \).

With this, we can approximate the expression in (14) by

\[
\bar{F}^{[M]}(\ell, s''_{\ell}) \approx \sum_{(i,j) \in B} Q_{ij} \sum_{s''} \bar{G}_{ij}(s''_{\ell}) \cdot \frac{1}{2N} \sum_{\ell \in T_N} \bar{F}^{[M]}(\ell, s'').
\]

(14)

where, for all \( s'' \in S'' \), the expression \( \frac{1}{2N} \sum_{\ell \in T_N} \bar{F}^{[M]}(\ell, s'') \) can be approximated by

\[
\frac{1}{2N} \sum_{\ell \in T_N} \left( \hat{b}''_{\ell+1} + \hat{y}''_{\ell+1} \right) \cdot \frac{1}{2N} \sum_{\ell \in T_N} \bar{F}^{[M]}(\ell, s'').
\]

(Clearly, the expression \( V(\hat{b}''_{\ell+1}|\hat{b}''_{\ell+M+1}, \hat{s}_{\ell}, \hat{y}_{\ell+1}^N) \) can be evaluated with the help of forward-backward algorithm type computations.)

I. Parallel Branches in the FSMS Model

So far, we have assumed that the FSMS trellis has no parallel branches. However, we can deal with parallel branches in the FSMS trellis in the same manner as outlined in Section IV.C of [1].

III. DISCUSSION OF THE RELEVANCE OF THE ABOVE CORRECTIONS TO OTHER PAPERS

(This section needs to be finalized.)

Section 5 of the paper [4] by Pfister uses some results of [1]. However, the setup is such that the trellis of the FSMS equals the trellis of the FSCM and such that the FSMS state uniquely determines FSM state. Therefore, the results in Section II-E are applicable and so no correction of the T-values is required.

ACKNOWLEDGMENT

We are very grateful to Wei Mao at the California Institute of Technology for pointing out a problem with the GBAA when trying to implement it for a certain FSMS/FSMC setup. The subsequent discussions then led to spotting the flaw in [1] and its suitable correction. Moreover, it is a pleasure to acknowledge discussions with Henry Pfister at Texas A&M University on the topic of this note.
REFERENCES


