ADAPTATION AND LEARNING IN MULTI-TASK DECISION SYSTEMS

Stefano Marano

DIEM, University of Salerno,
via Giovanni Paolo II 132,
I-84084 Fisciano (SA), Italy
marano@unisa.it

Ali H. Sayed

Ecole Polytechnique Federale de Lausanne (EPFL)
School of Engineering,
CH-1015 Lausanne, Switzerland
ali.sayed@epfl.ch

ABSTRACT

Adaptation and learning over multi-agent networks is a topic of great relevance with important implications. Elaborating on previous works on single-task networks engaged in decision problems, here we consider the multi-task version in the challenging scenario where the state of nature may change arbitrarily. We propose a data diffusion scheme for tracking these changes in real time, and investigate by numerical simulations the corresponding steady-state decision performance. For the slow-adaptation regime, the complete analytical characterization of the agents' status is provided, under the simplifying assumption that the network connection matrix is correctly estimated.

Index Terms— Adaptive networks, diffusion schemes, multi-task decisions, slow-adaptation regime.

1. INTRODUCTION

We consider a network of agents engaged in a decision task. Agents collect observations from the environment and diffuse their data through the network by local interactions with neighboring agents. Our focus is on multi-task networks in which agents are grouped into different clusters. Observations collected by agents belonging to the same cluster are independently drawn from the same statistical distribution, which is referred to as the state of nature for that cluster. Different clusters experience, in general, different states of nature. Each agent is aware of what the possible states of nature are, but does not know which cluster it belongs to. We consider the challenging scenario where the states of nature may change in an unpredictable and uncontrollable manner. Thus, the network is tasked to track these changes in real time, and the system design must manage the tradeoff between adaptation (tracking capability) and decision performance at steady-state (learning capability).

Multi-task decision problems are relevant to sensor network applications in heterogeneous environments. These applications include surveillance and tracking systems, social sensing, health monitoring, homeland security, disaster prevention and management, Earth observation, and many more — see [1–10] and the references therein.

The work of A. H. Sayed was supported in part by grant 205121-184999 from the Swiss National Science Foundation.

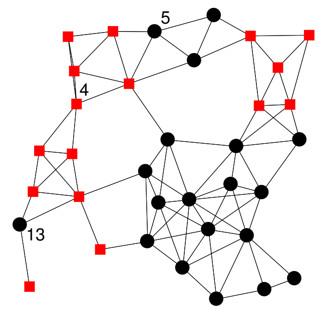

Fig. 1. An example of a multi-task network. Agents belonging to cluster 1, such as agents 5 and 13, are shown by black circles. Agents belonging to cluster 2, as agent 4, are shown by red boxes.

Our study can be cast in the line of research on multi-agent adaptive systems. The case of inference problems over single-task networks has been considered before, see, e.g., [11–18] for examples of consensus schemes and [1, 2, 19] for examples of diffusion schemes, with generalizations to multi-task networks in [3–5]. With specific reference to decision problems, the works [6–10] address the single-task setting, while we are not aware of studies on the multi-task decision problems as formalized in the next section.

2. MULTI-TASK DECISION FORMULATION

We consider a networked ensemble of S agents, grouped into clusters — see Fig. 1. The observations collected by agents belonging to a given cluster are drawn from a common distribution, selected from an assigned set of H probability mass functions (PMFs) $\{p_h(x)\}_{h=1}^{H}$, $x \in \mathcal{X}$, where $\mathcal{X}$ is a finite alphabet, common to all PMFs. The observation made by agent $k = 1, \ldots, S$, at time $i = 1, 2, \ldots$, is denoted by $\mathbf{x}_k(i)$. If $h \in \{1, \ldots, H\}$ is the state of nature at time i for the cluster that agent k belongs to, then $\mathbf{x}_k(i)$ is drawn from $p_h(x)$. Observations taken at different times and/or different agents are mutually independent. We make a distinction between neighbors $\mathcal{N}_k$ of agent k, namely those connected to k (by lines in Fig. 1), and *effective* neighbors $\mathcal{E}_k \subseteq \mathcal{N}_k$, which refer to neighbors belonging to the same cluster. For instance, agent 5 has 5 neighbors but only 3 effective neighbors. Note that, by definition, k is included in both $\mathcal{N}_k$ and $\mathcal{E}_k$.

Agents do not know their cluster. Thus, upon collecting the observation $\mathbf{x}_k(i)$ at time i, agent k computes H transformations as follows:

$$\mathbf{d}_k^{(h)}(i) \triangleq \log p_h(\mathbf{x}_k(i)), \quad h = 1, \dots, H, \tag{1}$$

and uses them to perform the updates:

$$\begin{cases} \mathbf{z}_k^{(h)}(i) = 0, & i = 0; \\ \mathbf{z}_k^{(h)}(i) = \mathbf{z}_k^{(h)}(i-1) + \mu[\mathbf{d}_k^{(h)}(i) - \mathbf{z}_k^{(h)}(i-1)], & i \geq 1. \end{cases} \tag{2}$$

The variables $\{\mathbf{z}_k^{(h)}(i)\}_{h=1}^H$ represent the status of the agent *isolated* from the network. According to the maximum likelihood approach, agent k can use these values to infer its cluster:

$$\widehat{\mathbf{h}}_k^{(\mathrm{loc})}(i) = \arg\max_h \mathbf{z}_k^{(h)}(i), \quad k = 1, \dots, S. \tag{3}$$

Agent k then communicates this *local decision* to its neighbors. In this way, at any time i, agent k is informed about the (random) set $\widehat{\boldsymbol{\mathcal{E}}}_k(i) \subseteq \mathcal{N}_k$ composed of the neighboring agents that believe they belong to the same cluster as agent k. Specifically,

$$\widehat{\boldsymbol{\mathcal{E}}}_k(i) \triangleq \{\ell \in \mathcal{N}_k \, : \, \widehat{\mathbf{h}}_\ell^{(\mathrm{loc})}(i-1) = \widehat{\mathbf{h}}_k^{(\mathrm{loc})}(i-1)\}, \tag{4}$$

where we note that the neighbors' local decisions at time $(i-1)$ are used to estimate the clustering at time i.

The sets $\{\widehat{\boldsymbol{\mathcal{E}}}_k(i)\}_{k=1}^S$ define the (random) right-stochastic nonnegative combination matrix $\boldsymbol{A}(i) = [\mathbf{a}_{k\ell}(i)]$ used in the diffusion algorithm to be introduced shortly, see (6). If $\widehat{\boldsymbol{\mathcal{E}}}_k(i) = \{k\}$, matrix $\boldsymbol{A}(i)$ contains $\mathbf{a}_{kk}(i) = 1$ at position (k,k) and all remaining entries over the k-th row are zero, which means that agent k does not combine its status with those of its neighbors. If, on the other hand, $|\widehat{\boldsymbol{\mathcal{E}}}_k(i)| > 1$, then we set

$$\mathbf{a}_{k\ell}(i) = \begin{cases} 0, & \ell \notin \widehat{\boldsymbol{\mathcal{E}}}_k(i), \\ \frac{1-a_k}{|\widehat{\boldsymbol{\mathcal{E}}}_k(i)|-1}, & \ell \in \widehat{\boldsymbol{\mathcal{E}}}_k(i) \setminus \{k\}, \\ a_k, & \ell = k. \end{cases} \tag{5}$$

The rationale is to assign the self-weight a_k to agent k itself, and constant weights to all other agents in $\widehat{\boldsymbol{\mathcal{E}}}_k(i)$. This reduces the design parameters appearing in the combination matrix to only the self-weight coefficients $\{a_k, k = 1, \dots, S\}$.

The ATC-type diffusion algorithm used in this work exploits matrix $\boldsymbol{A}(i)$ defined in (5) as follows [2]. For $h = 1, \dots, H$:

$$\begin{cases} \mathbf{w}_k^{(h)}(i) = 0, & i = 0, \\ \mathbf{v}_k^{(h)}(i) = \mathbf{w}_k^{(h)}(i-1) + \mu \left[\mathbf{d}_k^{(h)}(i) - \mathbf{w}_k^{(h)}(i-1)\right] & i \geq 1, \\ \mathbf{w}_k^{(h)}(i) = \sum_{\ell=1}^S \mathbf{a}_{k\ell}(i)\, \mathbf{v}_\ell^{(h)}(i), & i \geq 1, \end{cases} \tag{6}$$

and the decision made by agent k at time i is:

$$\widehat{\mathbf{h}}_k(i) = \arg\max_h \mathbf{w}_k^{(h)}(i). \tag{7}$$

Note that, over time $i = 1, 2, \dots$, two iterates are computed by any agent k, both driven by the same sequences $\{\mathbf{d}_k^{(h)}(i)\}_{h=1}^H$ defined in (1). The first, $\{\mathbf{z}_k^{(h)}(i)\}_{h=1}^H$ shown in (2) is used to define the matrix $\boldsymbol{A}(i)$, which is then exploited in the computation of the status $\{\mathbf{w}_k^{(h)}(i)\}_{h=1}^H$ in (6). Note also that, if the identity of the effective neighbors of all agents were known beforehand, then the connection matrix would reduce to a known time-invariant matrix; $\boldsymbol{A}(i) = A$.

Iterating (6) one obtains the explicit form:

$$\mathbf{w}_k^{(h)}(i) = \sum_{j=1}^{i} \mu(1-\mu)^{j-1} \times \sum_{\ell=1}^{S} \mathbf{b}_{k\ell}(i,j) \log p_h(\mathbf{x}_\ell(i-j+1)), \tag{8}$$

where we have introduced the matrix $\mathbf{B}(i,j) = [\mathbf{b}_{k\ell}(i,j)]$:

$$\mathbf{B}(i,j) \triangleq \prod_{m=i}^{i-j+1} \boldsymbol{A}(m), \quad j = 1, \dots, i, \text{ and } i = 1, 2, \dots \tag{9}$$

Note that (8) is the sum (over index j) of a random *triangular array* (see, e.g., [20]), because the argument of the outmost sum depends on i. Note also that the coefficients $\mathbf{b}_{k\ell}(i,j)$ depend in a nontrivial way on the the same observations $\mathbf{x}_k(i-j+1)$ appearing in the term $\log p_h(\mathbf{x}_k(i-j+1))$. For these reasons, in general, obtaining an exact statistical characterization of the agents' status $\mathbf{w}_k^{(h)}(i)$ is challenging. On the other hand, such characterization represents a fundamental building block for the design and analysis of practical detection algorithms. Accordingly, in the present work we derive an approximate and limiting characterization of the agents' status, which is then exploited to address the performance assessment of the decision algorithm (7) at steady-state. Our results can be the starting point for more accurate analysis of the system performance, which is left for future work.

3. STATISTICAL CHARACTERIZATION

The adaptive properties of ATC-like diffusion schemes are well-known, see e.g., [2]. Here, as done in similar studies [8–10], the focus is on the complementary aspect of the network learning capability, quantified by the decision performance at steady-state, assuming that the states of nature for all agents are constant for all times. This study will be conducted under the simplifying assumption that the diffusion algorithm employs the exact combination matrix A, namely we let $\widehat{\boldsymbol{\mathcal{E}}}_k(i) = \mathcal{E}_k$, as if the clustering operation would be made without errors. By substituting $\mathcal{E}_k$ for $\widehat{\boldsymbol{\mathcal{E}}}_k(i)$ in (5), we see that $\mathbf{B}(i,j) = B(j) = A^j$, see (9). Then, from (8):

$$\mathbf{w}_k^{(h)}(i) = \sqrt{\mu} \sum_{j=1}^{i} \mathbf{t}_k^{(h)}(i,j), \tag{10a}$$

$$\mathbf{t}_k^{(h)}(i,j) \triangleq \sqrt{\mu}(1-\mu)^{j-1} \sum_{\ell=1}^{S} b_{k\ell}(j) \log p_h(\mathbf{x}_\ell(i-j+1)). \tag{10b}$$

Likewise, for the local updates (2), letting $A = I_S$, where I_S denotes the $S \times S$ identity matrix, equation (10) yields:

$$\mathbf{z}_k^{(h)}(i) = \sqrt{\mu} \sum_{j=1}^{i} \mathbf{u}_k^{(h)}(i,j), \tag{11a}$$

$$\mathbf{u}_k^{(h)}(i,j) \triangleq \sqrt{\mu}(1-\mu)^{j-1} \log p_h(\mathbf{x}_k(i-j+1)). \tag{11b}$$

Let us arrange the quantities in (10b) and (11b) for $h = 1, \dots, H$, in a vector $\boldsymbol{q}_k(i,j)$ of length $2H$, as follows

$$\boldsymbol{q}_k(i,j) \triangleq \left[\mathbf{t}_k^{(1)}(i,j), \dots, \mathbf{t}_k^{(H)}(i,j), \mathbf{u}_k^{(1)}(i,j), \dots, \mathbf{u}_k^{(H)}(i,j)\right]^T, \tag{12}$$

and consider the triangular array of vectors:

$$\begin{array}{llllll} \boldsymbol{q}_k(1,1) & & & & & \\ \boldsymbol{q}_k(2,1) & \boldsymbol{q}_k(2,2) & & & & \\ \boldsymbol{q}_k(3,1) & \boldsymbol{q}_k(3,2) & \boldsymbol{q}_k(3,3) & & & \\ \dots & \dots & \dots & \dots & & \\ \boldsymbol{q}_k(i,1) & \boldsymbol{q}_k(i,2) & \boldsymbol{q}_k(i,3) & \dots & \boldsymbol{q}_k(i,i) & \\ \dots & \dots & \dots & \dots & \dots & \dots \end{array} \tag{13}$$

Let $\mathbb{E}$, $\mathbb{V}$, and $\mathbb{P}$ denote the expectation, variance, and probability operators, computed under the state of nature, say h^*, relative to the agent under consideration.

Theorem 1 *Suppose that all PMFs corresponding to the possible states of nature are strictly positive: $p_h(x) > 0$, $x \in \mathcal{X}$, $h = 1, \dots, H$. Then, in the limit $i \to \infty$ followed by $\mu \to 0$, the sum $\sum_{j=1}^{i} \left[\boldsymbol{q}_k(i,j) - \mathbb{E}\boldsymbol{q}_k(i,j)\right]$ of the zero-mean version of the i-th row of array (13), converges in distribution to a multivariate zero-mean Gaussian vector with covariance matrix Σ. The $2S \times 2S$ covariance matrix is given by*

$$\Sigma = \begin{pmatrix} \Lambda\,\beta_k(A) & \Lambda\,\gamma_k(A) \\ \Lambda\,\gamma_k(A) & \Lambda\,\frac{1}{2} \end{pmatrix}, \tag{14}$$

where the entries Λ_{mn} of the $S \times S$ matrix Λ are

$$\Lambda_{mn} = \mathbb{E}\Big[\big(\log p_m(\mathbf{x}) - \mathbb{E}\log p_m(\mathbf{x})\big) \big(\log p_n(\mathbf{x}) - \mathbb{E}\log p_n(\mathbf{x})\big)\Big], \tag{15}$$

and where

$$\beta_k(A) = \lim_{\mu\to 0} \sum_{j=1}^{\infty} \sum_{\ell=1}^{S} \mu(1-\mu)^{2j-2}\, b_{k\ell}^2(j), \tag{16}$$

$$\gamma_k(A) = \lim_{\mu\to 0} \sum_{j=1}^{\infty} \mu(1-\mu)^{2j-2}\, b_{kk}(j), \tag{17}$$

$$0 \le \gamma_k(A) \le \frac{1}{2}, \qquad \frac{1}{2S} \le \beta_k(A) \le \frac{1}{2}. \tag{18}$$

PROOF Fix k, i, and j, and note that each entry of the vector $\boldsymbol{q}_k(i,j)$ in (13) has finite variance because $p_h(x) > 0$ implies $|\log p_h(\mathbf{x})|^2 \le M$ for all h and some $M > 0$. Note also that the elements on each row of the array (13) are mutually independent. Let us denote by $\mathbb{COV}(\boldsymbol{q})$ the covariance matrix (computed under h^*) of the vector $\boldsymbol{q}$, and by $\mathbb{I}(\cdot)$ the indicator function. If, for some covariance matrix Σ and every $\epsilon > 0$,

$$\lim_{\mu\to 0} \lim_{i\to\infty} \sum_{j=1}^{i} \mathbb{COV}(\boldsymbol{q}_k(i,j)) = \Sigma, \tag{19}$$

$$\lim_{\mu\to 0} \lim_{i\to\infty} \sum_{j=1}^{i} \mathbb{E}\left[\|\boldsymbol{q}_k(i,j)\|^2\, \mathbb{I}\big(\|\boldsymbol{q}_k(i,j)\| > \epsilon\big)\right] = 0, \tag{20}$$

then $\sum_{j=1}^{i} \left[\boldsymbol{q}_k(i,j) - \mathbb{E}\boldsymbol{q}_k(i,j)\right]$ converges in distribution to a zero-mean Gaussian vector with covariance matrix Σ, by Lindeberg-Feller version of central limit theorem for triangular arrays of vectors [21, Prop. 2.27].

Consider the expression in (19). By exploiting the independence of the observations $\mathbf{x}_k(i)$ for different values of k, and assuming $m, n \in \{1, \dots, H\}$, simple calculations show that the covariance between the two variables $\mathbf{t}_k^{(m)}(i,j)$ and $\mathbf{t}_k^{(n)}(i,j)$ appearing in (10b), is given by

$$\mu(1-\mu)^{2j-2} \sum_{\ell=1}^{S} b_{k\ell}^2(j)\, \Lambda_{mn}. \tag{21}$$

By summing (21) for j that ranges from 1 to i, and taking the limit $i \to \infty$ followed by $\mu \to 0$, yields the $S \times S$ upper left corner of matrix Σ shown in (14), with $\beta_k(A)$ given in (16). Similar calculations hold for other values of (m,n), which shows that (19) is verified with Σ given by (14).

Consider next Lindeberg condition (20), and fix $\delta > 0$. By omitting the indexes (i,j) and the subscript k for notational simplicity, we have

$$\mathbb{E}\left[\|\boldsymbol{q}\|^2\, \mathbb{I}\big(\|\boldsymbol{q}\| > \epsilon\big)\right] \le \mathbb{E}\left[\frac{\|\boldsymbol{q}\|^{2+\delta}}{\epsilon^\delta}\, \mathbb{I}\big(\|\boldsymbol{q}\| > \epsilon\big)\right] \le \frac{\mathbb{E}\left[\|\boldsymbol{q}\|^{2+\delta}\right]}{\epsilon^\delta},$$

which shows that $\lim_{\mu\to 0} \lim_{i\to\infty} \sum_{j=1}^{i} \mathbb{E}\left[\|\boldsymbol{q}_k(i,j)\|^{2+\delta}\right] = 0$, known as Lyapunov-type condition [22], implies the limit in (20). Now, from (10)-(12),

$$\mathbb{E}\left[\|\boldsymbol{q}_k(i,j)\|^{2+\delta}\right] = \mu^{1+\frac{\delta}{2}} (1-\mu)^{(j-1)(2+\delta)} \times \mathbb{E}\left\{\sum_{h=1}^{H} \Big[\log p_h(\mathbf{x}_k(i-j+1)\Big]^2 + \sum_{h=1}^{H} \left[\sum_{\ell=1}^{S} b_{k\ell}(j) p_h(\mathbf{x}_\ell(i-j+1))\right]^2\right\}^{1+\frac{\delta}{2}} \tag{22a}$$

$$\le \mu^{1+\frac{\delta}{2}} (1-\mu)^{(j-1)(2+\delta)} \mathbb{E}\left\{HM + \sum_{h=1}^{H}\left[\sum_{\ell=1}^{S} b_{kl}^2(j) \times \sum_{\ell=1}^{S} \Big(\log_h(\mathbf{x}_\ell(i-j+1))\Big)^2\right]\right\}^{1+\frac{\delta}{2}} \tag{22b}$$

$$\le \mu^{1+\frac{\delta}{2}} (1-\mu)^{(j-1)(2+\delta)} (HM(1+S))^{1+\frac{\delta}{2}}, \tag{22c}$$

where (22b) follows by $|\log p_h(\mathbf{x})|^2 \le M$ and by Cauchy-Schwarz inequality. Inequality (22) shows that condition (20) is verified because $\lim_{\mu\to 0} \sum_{j=1}^{\infty} \mu^{1+\frac{\delta}{2}} (1-\mu)^{(j-1)(2+\delta)} = 0$. To conclude the proof, note that the inequalities (18) follow by recalling that the combination matrix A is nonnegative and right-stochastic and so are its powers $B(i) = A^i$, yielding

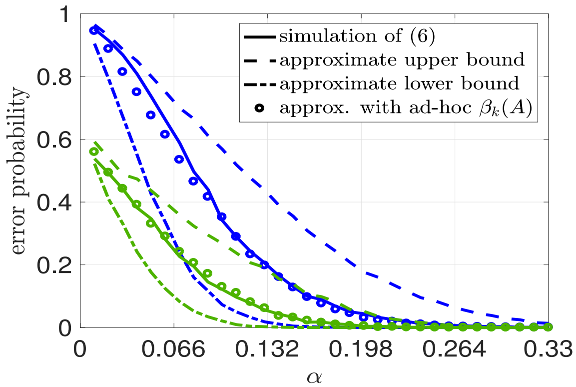

Fig. 2. Error probability of agent $k = 4$ for the network shown in Fig.1. Solid curves refer to simulations of the diffusion scheme (6), dashed curves to the upper bound, and dash-and-dotted curves to the lower bound. Small circles show an ad-hoc approximation obtained by adjusting the value of $\beta_k(A)$ used in deriving the lower bound. Different colors refer to different scenarios, see main text for details.

$\sum_{\ell=1}^S b_{k\ell}(i) = 1$, for all i. The bounds for $\gamma_k(A)$ follow immediately. The lower bound for $\beta_k(A)$ follows by Cauchy-Schwarz inequality $(\sum_{\ell=1}^S b_{k\ell}(i))^2 \leq S \sum_{\ell=1}^S b_{k\ell}^2(i)$. ■

Denoting by $\mathbf{w}_k^{(h)}$ and $\mathbf{z}_k^{(h)}$ the steady-state values of the algorithm outputs obtained by letting $i \to \infty$ in (10a) and (11a), Theorem 1 allows us to make the following approximation for $\mu \ll 1$ (with obvious notation):

$$\left[\mathbf{w}_k^{(1)}, \ldots, \mathbf{w}_k^{(H)}, \mathbf{z}_k^{(1)}, \ldots, \mathbf{z}_k^{(H)}\right]^T \sim \mathcal{N}(c, \mu\Sigma), \tag{23}$$

where $c = [\mathbb{E}\log p_1(\mathbf{x}), \ldots, \mathbb{E}\log p_H(\mathbf{x}), \mathbb{E}\log p_1(\mathbf{x}), \ldots \mathbb{E}\log p_H(\mathbf{x})]^T$, and Σ is given by (14). In particular,

$$\begin{aligned} \mathbf{w}_k^{(h)} &\sim \mathcal{N}\Big(\mathbb{E}\log p_h(\mathbf{x}),\ \mu\,\beta_k(A)\,\mathbb{V}\log p_h(\mathbf{x})\Big), \\ \mathbf{z}_k^{(h)} &\sim \mathcal{N}\Big(\mathbb{E}\log p_h(\mathbf{x}),\ \mu\,\tfrac{1}{2}\,\mathbb{V}\log p_h(\mathbf{x})\Big). \end{aligned} \tag{24}$$

Recalling that $1/(2S) \leq \beta_k(A) \leq 1/2$, from (24) we see that the beneficial effect of the collaboration among agents is quantified by the variance-reduction factor $\beta_k(A)$, which also encodes the topology of the network. We know that if $A = I_S$, then $\beta_k(A) = 1/2$ for all k, while if all the entries of A are equal to $1/S$, then $\beta_k(A) = 1/(2S)$ for all k. The former situation corresponds to isolated agents, while the latter corresponds to a fully connected single-task network with all agents belonging to the same cluster.

4. PERFORMANCE ASSESSMENT

Even under the simplifying assumption of knowing the connection matrix A, computing the error probability of agent k as shown in (7) requires difficult multi-dimensional integration of a jointly Gaussian distribution, which can only be done numerically. It is more convenient to exploit the theoretical results of the previous section to design a simplified setup for computer simulations, which achieves remarkable saving in terms of execution times with respect to the plain simulation of the diffusion algorithm (6). In particular, the statistical characterization of $\mathbf{z}_k^{(h)}$, for $h = 1, \ldots, H$, and $k = 1, \ldots, S$, provided in (23), is exploited to implement a standard Monte Carlo counting procedure for estimating $\mathbb{P}\{\arg\max_h \mathbf{z}_k^{(h)} \neq h^*\}$, see (3), where $h^\star$ is the state of nature relative to agent k. This probability is taken as an approximate upper bound of the actual performance. Likewise, exploiting (23), $\mathbb{P}\{\arg\max_h \mathbf{w}_k^{(h)} \neq h^*\}$ is estimated by Monte Carlo counting and is taken as an approximate lower bound of the actual performance, because of the assumption of knowing A.

With reference to the network shown in Fig. 1, composed of $S = 35$ agents, in Fig. 2 the performance bounds are compared to the results of computer simulations of the diffusion algorithm (6), for a scenario in which there are $H = 4$ possible states of nature, described by the following PMFs:

$$\begin{aligned} h = h_1 &\Rightarrow p_1 = [p_{11},\ p_{12},\ p_{13}], \\ h = h_2 &\Rightarrow p_2 = [p_{11} + \alpha,\ p_{12},\ p_{13} - \alpha], \\ h = h_3 &\Rightarrow p_3 = [p_{11} - \alpha,\ p_{12},\ p_{13} + \alpha], \\ h = h_4 &\Rightarrow p_4 = \left[p_{11} - \tfrac{\alpha}{2},\ p_{12} + \alpha,\ p_{13} - \tfrac{\alpha}{2}\right], \end{aligned} \tag{25}$$

wherein $p_{11} = p_{12} = p_{13} = 1/3$, and $0 < \alpha < 1/3$. Furthermore, we set $a_k = 0.5$ for all agents, see (5), $\mu = 0.05$, and $i = 1000$ algorithm steps are considered.

Figure 2 depicts the decision performance of agent $k = 4$, which belongs to cluster 2. The curves in blue in Fig. 2 refer to the case in which the state of nature for agents belonging to cluster 1 is $h = 4$, and that for agents belonging to cluster 2 is $h = 1$. The green curves refer to the case in which the state of nature for cluster 1 is $h = 3$, and that for cluster 2 is $h = 2$. The performance bounds are shown as dashed and dash-and-dotted curves. The small circles that provide a good approximation for the error probabilities were obtained by modifying the lower bound derived from $\mathbf{w}_k^{(h)}$ as follows. The variance-reduction factor $\beta_k(A)$, defined in (16), has been adjusted in an ad-hoc manner within the range of allowable values shown in (18), in order to closely fit the error probability of the diffusion scheme. This might suggest future approaches for obtaining reliable approximations, beyond the derived bounds.

It should be noted that the performance bounds are approximate. For instance, the upper bound obtained by $\mathbf{z}_k^{(h)}$ can be violated, because the error probability of agents having many neighbors but only a few effective neighbors, may be larger than the error probability of the agents themselves operating in isolation. This may happen, for instance, to agent 13 in Fig. 1. For these agents, our performance prediction is of limited utility, and a different approach should be pursued.

5. CONCLUSION

In this work, we developed an ATC-like diffusion scheme for multi-task networks engaged in decision tasks. Adaptation capabilities are ensured by design, and learning capabilities are investigated by computer simulations. A theoretical analysis is carried out to obtain approximate steady-state performance bounds, in the regime of small step-sizes. Insights are provided to derive approximations of the decision performance, rather than bounds, while performance prediction for agents with many non-effective neighbors remains an open problem.

6. REFERENCES

[1] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, “Diffusion strategies for adaptation and learning over networks,” *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, 2013.

[2] A. H. Sayed, “Adaptive networks,” *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[3] S.-Y. Tu and A. H. Sayed, “Distributed decision-making over adaptive networks,” *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1054–1069, Mar. 2014.

[4] J. Chen, C. Richard, and A. H. Sayed, “Multitask diffusion adaptation over networks,” *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4129–4144, Aug. 2014.

[5] S. Khawatmi, A. H. Sayed, and A. M. Zoubir, “Decentralized clustering and linking by networked agents,” *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3526–3537, July 2017.

[6] P. Braca, S. Marano, and V. Matta, “Enforcing consensus while monitoring the environment in wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, 2008.

[7] P. Braca, S. Marano, V. Matta, and P. Willett, “Asymptotic optimality of running consensus in testing binary hypotheses,” *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, 2010.

[8] V. Matta, P. Braca, S. Marano, and A. H. Sayed, “Diffusion-based adaptive distributed detection: Steady-state performance in the slow adaptation regime,” *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4710–4732, Aug. 2016.

[9] V. Matta, P. Braca, S. Marano, and A. H. Sayed, “Distributed detection over adaptive networks: Refined asymptotics and the role of connectivity,” *IEEE Trans. Signal and Inf. Process. over Networks*, vol. 2, no. 4, pp. 442–460, Dec. 2016.

[10] S. Marano and A. H. Sayed, “Detection under one-bit messaging over adaptive networks,” *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6519–6538, Oct. 2019.

[11] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[12] T. C. Aysal, M. J. Coates, and M. G. Rabbat, “Distributed average consensus with dithered quantization,” *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4905–4918, Oct. 2008.

[13] S. Kar and J. M. Moura, “Distributed consensus algorithms in sensor networks: quantized data and random link failures,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1383–1400, Mar. 2010.

[14] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, “Gossip algorithms for distributed signal processing,” *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[15] S. Kar and J. M. F. Moura, “Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 674–690, Aug. 2011.

[16] D. Bajovic, D. Jakovetić, J. Xavier, B. Sinopoli, and J. M. F. Moura, “Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4381–4396, Sept. 2011.

[17] S. Zhu, Y. C. Soh, and L. Xie, “Distributed parameter estimation with quantized communication via running average,” *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4634–4646, Sept. 2015.

[18] M. E. Chamie, J. Liu, and T. Basar, “Design and analysis of distributed averaging with quantized communication,” *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 3870–3884, Dec. 2016.

[19] F. S. Cattivelli and A. H. Sayed, “Distributed detection over adaptive networks using diffusion adaptation,” *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, 2011.

[20] W. Feller, *An Introduction to Probability and Its Applications, Volume 2*, John Wiley & Sons, New York, 1971.

[21] A. W. van der Vaart, *Asymptotic statistics*, New York: Cambridge University Press, 1998.

[22] P. Billingsley, *Probability and Measure*, Wiley-Interscience, New York, 3 edition, 1995.