Dynamic Quantization and Power Allocation for Multisensor Estimation of Hidden Markov Models

Nader Ghasemi, *Member, IEEE*, and Subhrakanti Dey, *Senior Member, IEEE*

***Abstract*—This paper investigates an optimal quantizer design problem for multisensor estimation of a hidden Markov model (HMMs) whose description depends on unknown parameters. The sensor measurements are simply binary quantized and transmitted to a remote fusion center over noisy flat fading wireless channels under an average sum transmit power constraint. The objective is to determine a set of optimal quantization thresholds and sensor transmit powers, called an optimal policy, which minimizes the long run average of a weighted combination of the expected state estimation error and sum transmit power. We analyze the problem by formulating an adaptive Markov decision process (MDP) problem. In this framework, adaptive optimal control policies are obtained using a nonstationary value iteration (NVI) scheme and are termed as *NVI-adaptive* policies. These NVI-adaptive policies are adapted to the HMM parameter estimates obtained via a strongly consistent maximum likelihood estimator. In particular, HMM parameter estimation is performed by a recursive expectation–maximization (EM) algorithm which computes estimates of the HMM parameters by maximizing a relative entropy information measure using the received quantized observations and the trajectory of the MDP. Under some regularity assumptions on the observation probability distributions and a geometric ergodicity condition on an extended Markov chain, the maximum-likelihood estimator is shown to be strongly consistent. It is shown that the NVI-adaptive policy based on this sequence of strongly consistent HMM parameter estimates is (asymptotically, under appropriate assumptions) average-optimal. Essentially, it minimizes the long run average cost of the weighted combination of the expected state estimation error and sum transmit power across the sensors for the HMM with true parameters in a time-asymptotic sense. The advantage of this scheme is that the policies are obtained recursively without the need to solve the Bellman equation at each time step, which can be computationally prohibitive. As is usual with value iteration schemes, practical implementation of the NVI-adaptive policy requires discretization of the state and action space, which results in some loss of optimality. Nevertheless, numerical results illustrate the asymptotic convergence properties of the parameter estimates and the asymptotically close to optimal performance of the adaptive MDP algorithm compared to the performance of an MDP based dynamic quantization and power allocation algorithm designed with perfect knowledge of the true parameters.**

***Index Terms*—Hidden Markov models (HMMs), maximum-likelihood (ML) estimation, quantization, state estimation, wireless sensor networks.**

Manuscript received February 20, 2011; revised February 21, 2011; accepted November 16, 2011. Date of publication December 09, 2011; date of current version June 22, 2012. This work was supported in part by the Australian Research Council (ARC) under Grant ARC DP 0985397. Recommended by Associate Editor L. Schenato.

The authors are with the Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, VIC 3010, Australia (e-mail: n.ghasemi@ee.unimelb.edu.au; nader.ghassemi@gmail.com; sdey@unimelb.edu.au; sdey@ee.unimelb.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2011.2179420

I. Introduction

WIRELESS sensor networks (WSNs) have attracted a significant level of research interest due to their wide range of current and potential applications such as in environment and structural health monitoring, surveillance, detection, and estimation of enemy targets in battlefield situations, and location aware services, etc. In detection/estimation tasks involving such WSNs, severe bandwidth constraints, limitations imposed by the fading wireless channels and the energy/power constraints of the small battery powered sensors have thrown up a new set of challenges. Various estimation problems with quantized (binary or with a small number of bits) data have been studied in [1] and [2]. Essentially, these studies consider parameter estimation problems. More recently, estimation of linear dynamical systems using quantized measurements have been considered in [3], while optimal quantizer design for Kalman filtering with quantized innovations has been investigated in [4].

In this paper, we study optimal quantization and power allocation for estimation of hidden Markov models (HMMs) with unknown parameters via multiple sensors communicating (binary) quantized measurements to a fusion center over fading wireless channels described by finite-state Markov chains. We present an adaptive Markov decision process (MDP) based approach for computing the optimal quantizer thresholds and the sensor transmit powers for minimizing expected state estimation error when the HMM parameters are unknown and need to be estimated from the quantized measurements.

A. Related Work

HMMs have long been considered as useful stochastic signal models in a broad range of areas, such as robotics, econometrics, biochemistry and biology. *Parameter* estimation of an HMM was first studied by Baum and Petrie [5] in 1966. They proposed a nonrecursive (offline) maximum-likelihood (ML) estimator for HMMs with observations taking only finitely many values. Assuming stationarity of the underlying Markov chain, they proved consistency and asymptotic normality of the ML estimator, where consistency refers to the almost sure convergence of the ML parameter estimates to the true parameter values under the probability measure induced by the true parameters, and asymptotic normality refers to the convergence in distribution of (a suitably scaled) parameter estimation error to a normal distribution. Since then there has been an abundant literature on HMM inference. Here, we briefly mention some related works. In [6], the conditions for consistency in [5] were weakened (strict positivity conditions on the Markov chain state transition probabilities and the state to observation probabilities were relaxed) and some identifiability results were presented. Later, Leroux in [7] proved consistency of the ML estimator for general HMMs under some mild conditions while its local

asymptotic normality was proved in [8]. *Recursive* (online) estimators for HMMs have been studied in [9] and [10] with their approaches being different in choosing scaling matrices used in the recursive procedure. In [9] and [10], no convergence results are provided, but their simulation studies show that the algorithms often converge well in practical cases. A batch recursive HMM estimator using a stochastic approximation algorithm has been proposed in [11] and [12] and proved to be consistent. Moreover, recursive estimation of parameters of an HMM defined as a Markov chain observed through a noisy infinite impulse response (IIR) channel has been studied in [13] with guaranteed convergence for some special cases. On line parameter estimation using recursive EM algorithm has been studied in [14] for estimation of various Markov-modulated time-series. In [15], a stochastic approximation algorithm for recursive estimation of HMMs has been proved to be consistent. These results have been generalized to autoregressive models with Markov regimes in [16] with an analysis of the asymptotic properties of the recursive algorithm.

Regarding *state* estimation of HMMs with quantized measurements, there are many studies reported in the literature which address state estimation of HMMs with various types of observations and under different constraints (see, e.g., [17]–[20]). In particular, a recent study [19] considered state estimation of a general HMM with binary quantized measurements sent over temporally correlated flat fading channels using an *unconstrained* MDP approach. The authors in [19] considered minimization of the long-term expected average of a cost function defined as the Lagrangian combination of expected state estimation error and total power consumption across the sensors. Furthermore, in a more recent study [20], the authors addressed the same problem as in [19] by using an alternative *constrained* MDP approach. In order to find optimal policies, they employed a linear programming technique which has a provably lower implementation complexity and is more efficient in terms of computations and memory requirements. In all of these studies, though, it is assumed that the parameters of the underlying HMM are *known* to the state estimation algorithm.

In most real applications, however, parameters of the HMM are unknown to the state estimator. This raises a new challenging problem involving joint estimation of the true parameter values and the HMM states based on the quantized measurements received over fading wireless channels under an average sum transmit power constraint, which is the primary focus of this paper.

B. Summary of Contributions

In contrast to related work such as [19] and [20], in this paper, we relax the restrictive assumption of true HMM parameter values being known, and explore the problem of *joint* state and parameter estimation of a general HMM, with binary quantized observations, whose description depends measurably on *unknown* parameters. Incorporation of a parameter estimation algorithm into our state estimation algorithm presented in [19] is the subject of the present research. We propose an approach based on a coupled adaptive MDP controller (which determines the optimal quantization thresholds and the optimal transmit powers at the sensors) and recursive EM based parameter estimator operating at the fusion center. This approach is described in detail below. Our approach (illustrated in Fig. 1) proposes a coupled algorithm in which state and parameter estimation are performed jointly as a single task at the fusion center. The first component of this task is an MDP module (also called the controller), whose function is to obtain an optimal policy (in our case, quantization thresholds and power levels) by performing state estimation so as to minimize expected state estimation error constrained on an average sum power (across all the sensors) budget. The MDP controller uses a nonstationary value iteration (NVI) scheme in order to obtain adaptive optimal policies. The NVI scheme *adapts* the optimal policy to the current estimate of unknown model parameters received from the second module, referred to as *Parameter Estimator*. The parameter estimator module finds the best model estimate by monitoring the sequence of MDP states and actions taken by the controller as well as the trajectory of the dynamical system via received observation sequences at the fusion center. The advantage of the nonstationary value iteration scheme is that the policies are obtained in an *iterative* manner without the need to solve the Bellman optimality equation at each time step, which in our case is highly computation-intensive.

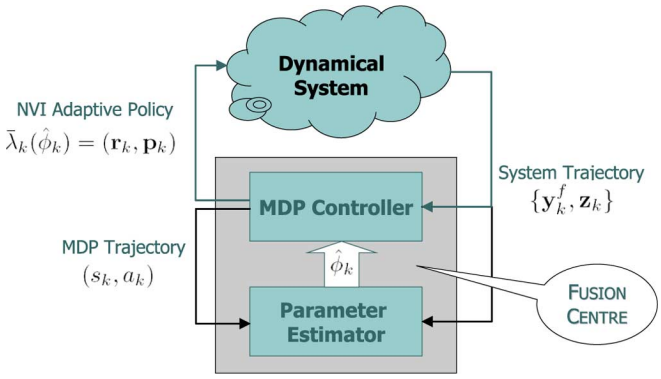


Fig. 1. Block diagram of the estimation setup for HMMs with unknown parameters using multiple sequences of binary-quantized observations.

Since all convergence results that can be obtained for the estimates to the true parameter value are asymptotic results, the optimality conditions will only be achieved in the "limit" sense. Therefore, we formulate our problem as an *infinite horizon* average cost *adaptive* MDP (essentially MDPs which depend on unknown parameters) problem, following the approach presented by Hernández-Lerma in [21] and further developed in [22]. There are many other approaches to average cost adaptive MDP problems, e.g., see the pioneering works of Kurano [23] and Mandl [24], and also [25] and [26]. The reason to choose an average cost criterion is because it depends only on the limiting behavior of the costs and not on the costs during the early periods [22]. Since the performance during the early stages does not contribute to the final average cost, the errors made by the controller module in the early steps when the parameter estimator module is still learning the parameters will have no effect in the limit. Therefore, if the parameter estimator module performs a strongly consistent parameter estimation, one can expect the controller performance to be average optimal. It is shown in [22] that this is indeed the case, under appropriate assumptions, which forms the basis of our coupled NVI-based adaptive MDP control algorithm and recursive EM-based parameter estimation algorithm.

Remark 1.1: Note that as is usual with any value iteration scheme associated with solving the Bellman equation for a long-run average cost MDP problem, numerical implementation of the NVI-based adaptive control algorithm requires a discretization of the state and action space. Thus, when solved with a discretized state and action space, there is some loss of optimality and the resulting NVI-adaptive policy only provides an approximation to the true "average-optimal" NVI-adaptive policy for the continuous state and action space. One expects that the approximation to get better as the number of discretized levels increases. Nevertheless, our numerical results illustrate that the performance of this discretized NVI-adaptive algorithm is close to that of the stationary policy obtained for the HMM with true parameters (albeit also with discretized state and action space), as presented in [19]. See also Remark 3.3 for comments on the relationship between the solution to the unconstrained weighted (based on a Lagrange multiplier associated with the average sum power) cost considered in this paper and the solution to the corresponding constrained MDP version of this problem, as considered (with known HMM parameters) in [20].

We use an online expectation-maximization (EM) algorithm based on the ML criterion for estimating the parameters of the HMM using the MDP trajectory and the observations received from the fusion center. We also establish strong consistency of the proposed online ML parameter estimators using the so-called mean ordinary differential equations method [27]. Numerical results are presented to illustrate the performance of our coupled dynamic quantization, sensor power allocation, and HMM parameter estimation algorithm. Note that a previous conference version of some of these results appeared in [28]. The current journal version provides an enhanced version with a detailed strong consistency proof of the online ML method and additional numerical results. Efforts have been made to reduce material already presented in [28] as much as possible without compromising the readability of the paper.

C. The Sequel

The material in this paper is organized as follows. In Section II, we present the model formulation for the dynamical system. Section III presents an MDP model for the problem formulated as an adaptive infinite-horizon average cost MDP problem. The optimal solution to the MDP problem is characterized by a *recursive* NVI scheme and corresponding adaptive policies. It is shown that, under appropriate assumptions, the policies adapted to the estimates of the model parameters are average optimal. In Section IV, we present the online EM algorithm based for HMM parameter estimation. Strong consistency of the proposed online ML estimator is established in Section V. Numerical results are given in Section VI. Section VII presents a summary of the paper with some concluding remarks. Detailed online estimation equations are provided in the Appendix.

II. Dynamical System Model

Notations: Throughout the paper, $\mathbb{R}$ and $\mathbb{N}$ denote the sets of real numbers and positive integers, respectively. We denote by C^n the class of n-times continuously differentiable functions. Also, $\mathbb{P}$ represents probability distributions with respect to some σ-finite measure. $\mathbb{E}_\phi$ stands for the expectation with respect to the probability measure $\mathbb{P}_\phi$, where $\mathbb{P}_\phi$ denotes a probability distribution parameterized by a parameter (vector) ϕ, with respect to some σ-finite measure. In this paper, vector means a column vector and $'$ denotes the transpose notation.

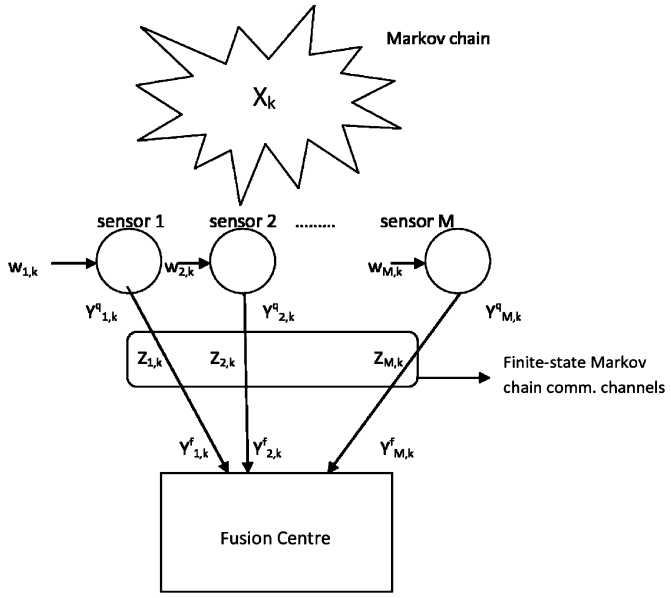


Fig. 2. Dynamical system model showing the Markov chain observed in noise by M sensors with their observations being quantized and sent to the Fusion Center via fading channels modeled by finite-state Markov chains.

A block-diagram of the multisensor observation setup of the dynamical system, including the communication channel over which the quantized sensor observations get sent to the fusion center is shown in Fig. 2. Below we describe this framework in details. The state process of the dynamical system evolves according to a discrete-time finite-state homogeneous first-order stationary Markov process $\{X_k\}_{k=1}^{\infty}$ with state space $\mathcal{S}(\phi) = \{\tilde{x}_1(\phi), \ldots, \tilde{x}_n(\phi)\}$ and transition probability matrix $\boldsymbol{\rho}(\phi) = [\varrho_{ij}(\phi)]$, where $\varrho_{ij}(\phi) = \mathbb{P}_\phi(X_k = \tilde{x}_j | X_{k-1} = \tilde{x}_i)$ for $i, j = 1, \ldots, n$. The order $n \in \mathbb{N}$ of the process $\{X_k\}$ is fixed and known, whereas, the state values $\tilde{x}_i$ are unknown. Let $\tilde{\mathbf{x}}_\phi = (\tilde{x}_1(\phi), \ldots, \tilde{x}_n(\phi))'$ denote the vector of unknown state values. The transition matrix $\boldsymbol{\rho}(\phi)$ and state space $\mathcal{S}(\phi)$ depend measurably on a parameter (vector) ϕ in a compact Euclidean space Φ. The "true" value of the parameter ϕ is denoted by $\phi^\circ \in \Phi$, and is assumed to be fixed but *unknown*. Note that for all $\phi \in \Phi$, we have $\varrho_{ij}(\phi) \geq 0$ and $\sum_j \varrho_{ij}(\phi) = 1$ for each i. The initial state probability vector of $\{X_k\}$ is denoted by $\pi = [\pi_i]$, where $\pi_i = \mathbb{P}(X_1 = \tilde{x}_i)$.

It is assumed that the Markov process $\{X_k\}$ is hidden and is observed indirectly by noisy measurements $Y_{m,k} = X_k + w_{m,k}$, $m = 1 \cdots M$, obtained from M sensors where M is fixed. Write $\mathbf{Y}_k = (Y_{1,k}, \ldots, Y_{M,k})'$ as the random vector of measurements obtained from the M number of sensors at time k. Also, let $\{\mathbf{Y}_k\}_{k=1}^{\infty}$ denote a vector of M random processes, with each process $\{Y_{m,k}\}_{k=1}^{\infty}$ being a sequence of conditionally independent random variables given a realization $\{x_k\}$ of $\{X_k\}$. Each random measurement $Y_{m,k}$ is characterized by a conditional density $f(\cdot|x_k; \theta^m(\phi))$ with respect to the Lebesgue measure $\mathcal{U}$ for $\theta^m : \Phi \mapsto \Theta$, where Θ is a Euclidean space. Write $\mathbf{w}_k = (w_{1,k}, \ldots, w_{M,k})'$ and let $\{\mathbf{w}_k\}_{k=1}^{\infty}$ be a vector of M independent noise processes, where each process $\{w_{m,k}\}_{k=1}^{\infty}$ is assumed to be an independent and

identically distributed (i.i.d.) sequence of scalar real-valued innovations with known marginal distribution parameterized by a vector $\theta^m(\phi) \in \Theta$.

Remark 2.1: In our numerical studies, we shall deal with zero-mean white Gaussian noise processes $\{\mathbf{w}_k\}_{k=1}^{\infty}$, thus having $f(\cdot|x_k;\theta^m(\phi)) \sim \mathcal{N}(y;x_k,\theta^m=\sigma_m(\phi))$, where we use the notation $Y_{m,k} \sim \mathcal{N}(y;\mu,\sigma)$ as a short-hand for the univariate Normal probability density function, where $\mathcal{N}(y;\mu,\sigma) = (2\pi\sigma^2)^{-1/2}\exp[-0.5(y-\mu)^2\sigma^{-2}]$.

Due to severe bandwidth limitations in sensor networks, the measurements $\mathbf{Y}_k$ are then quantized according to a threshold-based binary quantization scheme with the sequence $\{\mathbf{r}_k\}_{k=1}^{\infty} \triangleq \{(r_{1,k},\ldots,r_{M,k})\}_{k=1}^{\infty}$ denoting the sequence of quantization thresholds. Note that the analysis in this paper can be readily extended to consider higher number of quantizer thresholds, albeit at the expense of increased computational complexity. Let $\mathbf{Y}_k^q = (Y_{1,k}^q,\ldots,Y_{M,k}^q)'$ represent the quantized data at time k, where $Y_{i,k}^q = b_1$ if $Y_{i,k} < r_{i,k}$ and $Y_{i,k}^q = b_2$ otherwise. The mth sensor transmits its quantized output $Y_{m,k}^q$, with power level $p_{m,k}$ to a remote fusion center over a discrete time flat fading channel. Let $\{\mathbf{p}_k\}_{k=1}^{\infty} \triangleq \{(p_{1,k},\ldots,p_{M,k})\}_{k=1}^{\infty}$ be the sequence of power levels and $\mathbf{Z}_k \triangleq (Z_{1,k},\ldots,Z_{M,k})'$ be the sensors' channel state vector at time k. We model each channel state process $\{Z_{m,k}\}_{k=1}^{\infty}$ as a stationary ergodic Markov chain with state space $\mathbb{C} = \{\tilde{c}_1,\ldots,\tilde{c}_u\}$ and transition probability matrix[1] $\mathbf{C}^m = [c_{ij}^m]$, where $c_{ij}^m = \mathbb{P}(Z_{m,k} = \tilde{c}_j | Z_{m,k-1} = \tilde{c}_i)$, $1 \le i,j \le u$. Each channel state $\tilde{c}_i$ may represent a value of the channel gain. The initial state distribution of $\{Z_{m,k}\}$ is given by $\pi^m = [\pi_i^m]$, where $\pi_i^m = \mathbb{P}(Z_{m,1} = \tilde{c}_i)$.

Remark 2.2: Note that finite-state Markov chain models have often been used in the information theory literature to characterize wireless channels. The channel is typically modeled by appropriately partitioning the range of the received signal-to-noise ratio (SNR) into a set of intervals (states) using a suitable partitioning criteria, e.g., for the Gilbert–Elliot model for a two-state channel see [29], [30]. For Markov models with higher number of states see [31] and references therein.

We assume that the channel state information is perfectly known at the receiver which also knows the transition probabilities of the Markov fading channel. Note that such channel estimation can be carried out in a training phase periodically once every fading block using pilot symbols from the fusion center under the condition that the channels between each sensor and the fusion center is symmetric (i.e., the sensor to fusion center channel and the fusion center to sensor channel are identical such as in a time division duplex (TDD) scheme). This is a fair assumption since the fusion center is typically not limited by energy/power constraints. The transmission power for each sensor is chosen from a set $\mathbb{V}$ of finitely many discrete power levels, which is generally the case for most practical sensor systems.

Let $\mathbf{Y}_k^f = (Y_{1,k}^f,\ldots,Y_{M,k}^f)'$ be the vector of decoded binary symbols at the fusion center, where $Y_{m,k}^f \in \{b_1,b_2\}$ as well. $Y_{m,k}^f$ is described by the following channel input–output transition probability $q_{ij}^m(\tilde{c},\tilde{p}) \triangleq \mathbb{P}(Y_{m,k}^f = b_j | Y_{m,k}^q = b_i, Z_{m,k} = \tilde{c}, p_{m,k} = \tilde{p})$, where $i,j \in \{1,2\}$, $\tilde{c} \in \mathbb{C}$, $\tilde{p} \in \mathbb{V}$. The off-diagonal entries in the input–output transition matrix $\mathbf{Q}^m(\tilde{c},\tilde{p}) \triangleq [q_{ij}^m(\tilde{c},\tilde{p})]$ are called crossover probabilities. A graphical illustration of a two-state Markov chain communication channel $Z_{m,k}$ for the mth sensor is given in Fig. 3, where the binary input–output transition probabilities from $Y_{m,k}^q$ to $Y_{m,k}^f$ are shown as a function of the channel states $\tilde{c}_1$ and $\tilde{c}_2$. We assume that the sensors use a simple binary phase shift keying (BPSK) modulation scheme to transmit the binary quantized measurements over orthogonal additive white Gaussian noise channels. The crossover probability can be computed as $\boldsymbol{\epsilon}_k^m = Q(\sqrt{\gamma g_{m,k}^2 p_{m,k} \sigma_v^{-2} d_m^{-\varsigma}})$, where γ is a constant, $g_{m,k}$ is gain of the wireless channel, σ_v^2 is the variance of the additive white Gaussian channel noise, d_m is the distance between the mth sensor and the fusion center, and ς is the path loss exponent of the wireless channel, and $Q(\cdot)$ is the complementary standard normal cumulative distribution function (cdf). Note that under certain standard symmetry assumptions on the modulation scheme and noise, the channel input–output transition probability matrix becomes a symmetric matrix (the channel is called a binary symmetric channel (BSC)), which is the case assumed in our analysis for simplicity. For further details on how to compute the crossover probabilities, see [19].

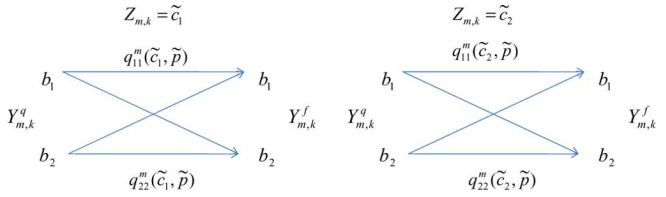


Fig. 3. Illustration of a two-state Markov chain fading communication channel.

We may now specify an HMM corresponding to the observation sequence $\{\mathbf{Y}_k^f\}_{k=1}^{\infty}$, decoded at the fusion center, by $\mathcal{H} = (\boldsymbol{\rho}(\phi),\mathcal{S}(\phi),\pi,\boldsymbol{\Psi}(\theta(\phi)))$, where $\theta(\phi) = (\theta^1(\phi),\ldots,\theta^M(\phi))'$, and $\boldsymbol{\Psi}$, the so-called state-to-observation[2] probability matrix, is a diagonal matrix with ith diagonal entry $\psi_i(\mathbf{y}_k^f,\mathbf{z}_k,\mathbf{r}_k,\mathbf{p}_k;\theta(\phi))$, $i=1,\ldots,n$ being the conditional probability mass function of $\mathbf{Y}_k^f$ defined as $\mathbb{P}_\phi(\mathbf{Y}_k^f = \mathbf{y}_k^f | X_k = \tilde{x}_i, \mathbf{r}_k, \mathbf{Z}_k = \mathbf{z}_k, \mathbf{p}_k; \theta(\phi))$.

The task of the fusion center is to find the optimal quantizer thresholds $\mathbf{r}_k$ and the optimal sensor transmit powers $\mathbf{p}_k$ while jointly estimating the state and the parameters of the underlying Markov chain $\{X_k\}$ with the objective being minimization of average state estimation error subject to an average sum power constraint across the sensors. The optimal quantizer thresholds and the optimal transmit powers for each sensor are then fed back to the individual sensors via the fusion center to sensor feedback channels (assumed to be delay and error free) to be used for sensor transmissions at the next time slot, see [19] for further details.

Definition 2.1: [28] Define the information state vector $\tilde{\alpha}_{k+1;\hat{\phi}^{(k)}}$ with ith element $\tilde{\alpha}_{k+1;\hat{\phi}^{(k)}}(i)$, also known as normalized HMM filter density or normalized forward variable, being defined as $\mathbb{P}_\phi(X_{k+1} = \tilde{x}_i | \mathcal{D}_{k+1}, \mathcal{B}_{k+1}; \hat{\phi}^{(k)})$, where $\hat{\phi}^{(k)} = (\hat{\boldsymbol{\rho}}^{(k)}, \hat{\mathcal{S}}^{(k)}, \hat{\boldsymbol{\theta}}^{(k)})$ denote the sequence of estimates of model parameters up to time k, and $\mathcal{D}_k$, $\mathcal{B}_k$ are the σ-fields generated by $(\mathbf{Y}_l^f, \mathbf{Z}_l)$, $(\mathbf{r}_l, \mathbf{p}_l)$, $l \le k$, respectively. Also, define the filtered state estimate as $\hat{X}_{k+1;\hat{\phi}^{(k)}} \triangleq \mathbb{E}_\phi[X_{k+1} | \mathcal{D}_{k+1}, \mathcal{B}_{k+1}; \hat{\phi}^{(k)}]$.

[1]To simplify our subsequent analysis, we assume that $c_{ij}^m > 0$ for all $1 \le i, j \le u$ and $1 \le m \le M$.

[2]see [19] for details on deriving the state-to-observation probabilities.

The following recursive equations for computing the conditional filtered probability densities are well known and we state them without proof:

Lemma 2.1: $\tilde{\alpha}_{k+1;\hat{\phi}^{(k)}}$ and $\hat{X}_{k+1;\hat{\phi}^{(k)}}$ may be computed inductively using the following forward recursion:

$$\bar{\alpha}_1 = \mathbf{\Psi}(\mathbf{y}_1^f, \mathbf{z}_1, \mathbf{r}_1, \mathbf{p}_1; \hat{\theta}_0(\hat{\phi}_0))\, \pi;$$
$$\bar{\alpha}_{k+1} = \mathbf{\Psi}(\mathbf{y}_{k+1}^f, \mathbf{z}_{k+1}, \mathbf{r}_{k+1}, \mathbf{p}_{k+1}; \hat{\theta}_k)\, \boldsymbol{\rho}' \bar{\alpha}_k,\ k \geq 1;$$
$$\tilde{\alpha}_{k+1;\hat{\phi}^{(k)}} = \langle \bar{\alpha}_{k+1}, \mathbb{1}_n \rangle^{-1}\, \bar{\alpha}_{k+1},\ k \geq 0;$$
$$\hat{X}_{k+1;\hat{\phi}^{(k)}} = \langle \tilde{\boldsymbol{\rho}}_{\hat{\phi}_k}, \tilde{\alpha}_{k+1;\hat{\phi}^{(k)}} \rangle;$$

where $\mathbb{1}_n$ is the n-dimensional column vector with all elements equal to one and $\bar{\alpha}_{k+1}$ is the (unnormalized) forward variable with respect to the Lebesgue measure $\mho$.

The state space of the forward variable $\tilde{\alpha}_{k+1;\hat{\phi}^{(k)}}$ can be defined as the simplex $\mathbb{T}_{\tilde{\alpha}} = \{\tilde{\alpha} \in \mathbb{R}_+^n \mid \langle \tilde{\alpha}, \mathbb{1}_n \rangle = 1\} \subset \mathbb{R}^n$. For numerical tractability of solving an associated Bellman equation later, we approximate the continuum information state $\tilde{\alpha}_{k+1;\hat{\phi}^{(k)}}$ by a discretized vector $\alpha_{k+1;\hat{\phi}^{(k)}} \in \mathbb{T}$, where $\mathbb{T} \subset \mathbb{T}_{\tilde{\alpha}}$ is the state space of the discretized forward variable $\alpha_{k+1;\hat{\phi}^{(k)}}$. Note that after discretization, it is ensured that $\langle \alpha, \mathbb{1}_n \rangle = 1, \forall\ \alpha \in \mathbb{T}$ (see [17] for further details). Let $\mathbb{U}$ denote the range space of each variable $r_{m,k}$ which is, in general, equal to $\mathbb{R}$ for $\forall\ m \in \{1 \cdots M\}$. The theory in Section III holds true, in general, for $\mathbb{U} = \mathbb{R}$ with the usual topology. However, in further analysis, to simplify the implementation of our value iteration algorithm, we restrict the action space $\mathbb{U}$ to a finite set of discrete values in $\mathbb{R}$. For further details on this discretization procedure, again see [17].

III. MDP Controller

A. Preliminaries

We define $(\mathbb{S}, \mathbb{A}, g(\cdot;\phi), p(\cdot|\cdot;\phi))$ to be an *adaptive* MDP depending on an unknown parameter $\phi \in \Phi$, where $\mathbb{S}$ and $\mathbb{A}$ are called state and action spaces, respectively. $\mathbb{S}$ is assumed to be a nonempty countable (possibly infinite) set endowed with a discrete topology. The action space $\mathbb{A}$ is assumed to be a nonempty Borel space defined as a Borel subset of a complete separable metric space. Further, $g(\cdot;\phi)$, the so-called immediate (or per-stage) cost function, is a continuous measurable function on $\mathbb{K} \times \Phi$ and $p(\cdot|\cdot;\phi) \in \mathbb{Q}$ is the transition law of the MDP. $\mathbb{K}$ is the set of admissible state-action pairs defined as $\mathbb{K} \triangleq \{(s,a) | s \in \mathbb{S}, a \in \mathbb{A}(s)\}$ which is a topological subspace of $\mathbb{S} \times \mathbb{A}$.

Remark 3.1: Note that $g(\cdot;\phi)$ and $p(\cdot|\cdot;\phi)$ depend measurably on a parameter (vector) ϕ whose "true" value, ϕ°, is fixed but unknown. It is also assumed that the set of admissible parameter values is given by a Borel space Φ. Henceforth, for simplicity, we may use shorter notations $g(\phi)$ and $p(\phi)$.

Remark 3.2: In order to simplify the implementation of our optimal quantization and power allocation algorithm, we assume the action space $\mathbb{A}$ to be a finite topological space endowed with the discrete topology. However, the theoretical results presented in the following hold in greater generality for all action sets $\mathbb{A}$ which are Borel spaces. In particular, $\mathbb{A}$ can be a countable (possibly infinite) set endowed with the discrete topology, or a compact metric space which is complete and separable, or in general $\mathbb{R}^d$, $d \in \mathbb{N}$ endowed with the usual topology [22].

B. Adaptive Markov Decision Process Model

In this section, we formulate our quantization problem as an adaptive infinite-horizon average cost MDP problem. Let $\mathcal{M}_\phi = (\mathbb{S}, \mathbb{A}, g(\phi), p(\phi))$ be an adaptive MDP depending on a parameter vector $\phi \in \Phi$ as defined above, where $\mathbb{S} = \mathbb{T} \times \mathbb{C}^M$, $\mathbb{A} = \mathbb{U}^M \times \mathbb{V}^M$ are corresponding state and action spaces, respectively. The immediate cost function $g(\phi)$ is defined by $g(\alpha_{k;\phi}, \mathbf{z}_k, \mathbf{r}_k, \mathbf{p}_k; \phi) \triangleq \mathcal{E}(\alpha_{k;\phi}) + \beta \mathcal{Q}(\mathbf{p}_k)$, which is a weighted combination of two cost functions: the conditional expected state estimation error $\mathcal{E}(\alpha_{k;\phi}) \triangleq \mathbb{E}_\phi[\ |X_k - \hat{X}_{k;\phi}|^2\ |\mathcal{D}_k, \mathcal{B}_k; \phi] = \sum_{i=1}^n [\tilde{x}_i(\phi) - \sum_{j=1}^n \tilde{x}_j(\phi) \alpha_{k;\phi}(j)]^2 \alpha_{k;\phi}(i)$, and the total power consumption across the sensors at time k, $\mathcal{Q}(\mathbf{p}_k) \triangleq \sum_{m=1}^M p_{m,k}$, $p_{m,k} \in \mathbb{V}$. Note that here we consider a constraint on the total average power consumption across all sensors, which is motivated by possible scenarios where the total power consumption may be limited due to a minimum lifetime constraint on the network, or where fair performance comparison irrespective of the number of sensors may be required. In a clustered sensor network scenarios, total power consumption within a cluster may also be limited to reduce interference on a neighboring cluster [32]. It is of course possible to consider individual power constraints in our framework at the cost of increased complexity due to the introduction of a Lagrange multiplier for each sensor's average power constraint. In the special case of all sensors having identical signal-to-noise ratio and statistically identical fading channels and identical average power constraints of P_{av}, considering a total power constraint of MP_{av} would yield the same optimal solutions as in the case with individual power constraints.

Remark 3.3: Note that in the combined cost function $g(\phi)$ by using the weighting factor β, a constrained quantization problem is transformed into an unconstrained (Lagrangian) problem. The Lagrangian is the combination of the original cost (mean square estimation error) and the average sum power constraint (e.g., $\mathbb{E}[\sum_{m=1}^M p_{m,k}] \leq P_{av}$]) weighted by some constant factor $\beta \geq 0$ called the Lagrange multiplier or tradeoff parameter. A suitably chosen value of β ensures that the required sum power constraint is satisfied with equality. As opposed to duality theory in convex optimization, the analysis of a duality gap or conditions under which strong duality holds in the case of a constrained MDP problem is far more complicated, especially in the case of state and actions taking values in Borel spaces, which is the case in our paper. This is due to the conversion of the partially observed MDP to a fully observed MDP via the information state approach and also the fact that the sensor quantizer thresholds and transmission powers can take real values. One can impose a compactness assumption on the action space and the information state lives on the simplex and is thus compact. Strong duality between the constrained MDP and the corresponding unconstrained MDP for compact Borel state and action spaces has been shown to hold via equivalent linear program formulations under some conditions in [33]. For more general Borel state and action spaces and unbounded costs, similar dual linear programming formulations have been used to prove strong duality under certain assumptions in [34]. Most of the conditions provided in [33] can be verified to hold for our problem once the maximum sensor transmit powers are constrained, due to continuity and

boundedness of the cost and constraint functions and weak continuity of the Markov transition kernel for a given state and action value. However, some of the additional conditions on when a deterministic stationary policy as a solution to the dual problem also solves the constrained problem are harder to verify. Even if these conditions can be verified, note that due to discretization, there will always be some loss in optimality, as explained in Remark 1.1 earlier. Hence from a practical implementation point of view, the NVI-adaptive policy for the Lagrange parameter based cost will always have a duality gap with the optimal policy for the constrained problem.

Remark 3.4: Due to the lack of exact knowledge of the Markov process state X_k at the fusion center, direct minimization of the mean square state estimation error $\mathbb{E}|X_k - \hat{X}_{k;\phi}|^2$ would require solving a partially observable MDP (POMDP) problem. In order to avoid the complications in determining the solution to a POMDP problem, we express the error cost function $\mathcal{E}(\cdot)$ in terms of the information state variable α_k in order to convert the partially observable MDP problem into a fully observable MDP problem, see [17], [19]. This technique of converting a POMDP to a fully observed MDP in terms of the information state model is standard. Note however that due to this conversion, we end up with a Borel state space MDP, which is usually approximated by a finite state MDP via suitable discretization of the information state space.

For a given parameter value $\phi \in \Phi$, if the MDP $\mathcal{M}_\phi$ is in state $s = (\alpha, \mathbf{z}) \in \mathbb{S}$ and action $a = (\mathbf{r}, \mathbf{p}) \in \mathbb{A}(s)$ is taken, then the observation $\mathbf{y}^f$ will be received at the fusion center and the MDP state changes to $\acute{s} = (\acute{\alpha}, \acute{\mathbf{z}}) \in \mathbb{S}$ according to the transition probability distribution $p(\acute{s}|s, a; \phi)$ which is computed by $p_z(\acute{\mathbf{z}} \mid \mathbf{z}) \left\langle \boldsymbol{\Psi}(\mathbf{y}^f, \acute{\mathbf{z}}, \mathbf{r}, \mathbf{p}; \theta(\phi)) \boldsymbol{\rho}'_\phi \alpha,\ \mathbb{1}_n \right\rangle$, where for $\mathbf{z} = (\tilde{c}_{i_1}, \ldots, \tilde{c}_{i_M})'$ and $\acute{\mathbf{z}} = (\tilde{c}_{j_1}, \ldots, \tilde{c}_{j_M})'$, $p_z(\acute{\mathbf{z}} \mid \mathbf{z})$ is the product of M channel transition probabilities computed by $\prod_{m=1}^{M} c^m_{i_m j_m}$ for $i_m, j_m \in \{1, \ldots, u\}$. From Lemma 2.1, it is clear that the value of the forward variable $\acute{\alpha}$ in the next MDP state $\acute{s}$ is obtained by recursion $\left[\langle \bar{\alpha},\ \mathbb{1}_n \rangle^{-1}\ \bar{\alpha}\right]_{round}$, where $[\cdot]_{round} : \mathbb{T}_{\tilde{\alpha}} \mapsto \mathbb{T}$ is the discretization operator (that rounds the argument to the nearest discretized value) for the information state as described in [17], and $\bar{\alpha}$ is the unnormalized forward variable with respect to the Lebesgue measure $\mho$ computed by $\boldsymbol{\Psi}(\mathbf{y}^f, \acute{\mathbf{z}}, \mathbf{r}, \mathbf{p}; \theta(\phi))\ \boldsymbol{\rho}'_\phi \alpha$.

For each (fixed) value of $\phi \in \Phi$, we specify an objective function $J^\lambda_\phi(\mathring{s})$, expressed as the long-term average expected cost per time step, or simply the average cost defined by

$$J^\lambda_\phi(\mathring{s}) \triangleq \limsup_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}^\lambda_{\mathring{s},\phi} \times \left[g(\alpha_{k;\phi}, \mathbf{z}_k, \mathbf{r}_k, \mathbf{p}_k; \phi) \big| \boldsymbol{\alpha}_1 = \mathring{\boldsymbol{\alpha}}, \mathbf{Z}_0 = \mathring{\mathbf{z}} \right] \quad (1)$$

where $\mathring{s} = (\mathring{\boldsymbol{\alpha}}, \mathring{\mathbf{z}})$ is the initial condition, $\lambda = (\mathbf{r}, \mathbf{p}) \in \boldsymbol{\Lambda}$ is a policy, and $\boldsymbol{\Lambda}$ is the non-empty space of all admissible randomized history-dependent policies. For each fixed value of $\phi \in \Phi$, $\mathbb{E}^\lambda_{\mathring{s},\phi}$ denotes the expectation with respect to $\mathbb{P}^\lambda_{\mathring{s},\phi}$ which is the unique probability measure on $(\Omega, \mathcal{F})$ for a given policy $\lambda \in \boldsymbol{\Lambda}$, and initial state $\mathring{s} \in \mathbb{S}$. The function $J^\lambda_\phi(\mathring{s})$ is a performance metric for our quantization problem measuring the performance when a given policy λ is used and the system starts with the initial condition $\mathring{s}$.

Our quantization problem may then be expressed as an adaptive stochastic control problem defined as follows. Determine an average-optimal policy $\lambda^*_{\phi^\circ}$ and its corresponding average optimal cost $J^*_{\phi^\circ}$ as the solution to the following optimization problem

$$(\mathbf{P}) : \inf_{\lambda \in \boldsymbol{\Lambda}} J^\lambda_\phi(s), \text{ for } \forall s \in \mathbb{S}$$

where J^λ_ϕ is the function defined in (1) and $\phi \in \{\hat{\phi}_k\}$ is a strongly consistent convergent sequence of estimates for ϕ°. The idea behind this is to compute a sequence of estimates $\{\hat{\phi}_k\}_{k=1}^\infty$ of the true parameter ϕ°, and show that if $\hat{\phi}_k$ converges to ϕ° $\mathbb{P}^\lambda_{s,\phi^\circ}$-a.s. as $k \to \infty$ then policies (suitably[3]) adapted to the *approximating* MDP sequence $\mathcal{M}_{\hat{\phi}_k} = (\mathbb{S}, \mathbb{A}, g(\hat{\phi}_k), p(\hat{\phi}_k))$ are average-optimal for the true MDP $\mathcal{M}_{\phi^\circ}$ [22]. We develop this approach in Section IV. We introduce the following assumptions from [28].

Assumptions 3.1: For the adaptive MDP $\mathcal{M}_\phi$ the following hold:

A1) Each state $s \in \mathbb{S}$ is associated with a nonempty measurable compact set $\mathbb{A}(s) \subseteq \mathbb{A}$ of admissible actions when the MDP is in state s.

A2) The immediate cost $g(s, a; \phi)$ is a continuous function of $a \in \mathbb{A}(s)$ for $\forall s \in \mathbb{S}$ uniformly in $\phi \in \Phi$.

A3) For some constant G, the immediate cost function g satisfies $|g(s, a; \phi)| \leq G < \infty$ uniformly in ϕ.

Remark 3.5: Note that for the adaptive MDP $\mathcal{M}_\phi$ defined in Section III-B, Assumptions **A1)** and **A3)** in 3.1 are satisfied due to the fact that $\mathbb{A}(s) = \mathbb{A}$ for $\forall s \in \mathbb{S}$, and $\mathbb{A}$ and $\mathbb{S}$ are finite topological spaces. It can be shown that the immediate cost function g is a continuous function of the action of the power level $\mathbf{p}_k$, and quantization thresholds $\mathbf{r}_k$ for the choice of our measurement noise distribution (Gaussian) and the continuous dependence of the channel crossover probabilities on the power levels. Thus, **A2)** in Assumption 3.1 is also satisfied. Moreover, the parameter space Φ is assumed to be compact and its elements are admissible parameter vectors which satisfy all the required constraints such as those imposed by transition probabilities of the Markov process $\{X_k\}_{k=1}^\infty$.

C. ϕ-Optimality Equation

It has been shown (cf. [35, Th. 5.5.3]) that the search domain for optimal policies in the stochastic control problem $(\mathbf{P})$ may be restricted only to the space of Markov policies instead of the general domain $\boldsymbol{\Lambda}$ of randomized history-dependent policies. Let $\mathbb{F}$ denote the space of all deterministic Markov decision rules defined as measurable functions $\lambda_\phi(s) : \mathbb{S} \times \Phi \mapsto \mathbb{A}$ such that $\lambda_\phi(s) \in \mathbb{A}(s)$ for every $s \in \mathbb{S}$ and $\phi \in \Phi$. Assume that for each $\phi \in \Phi$, the action $a_k = (\mathbf{r}_k, \mathbf{p}_k)$ at each time step k is determined by a stationary deterministic Markov policy $\lambda^\infty_\phi = \{\lambda_\phi, \lambda_\phi, \ldots\}$, where $a_k = \lambda_\phi(s_k)$. Henceforth, for brevity, λ^∞_ϕ may be denoted by λ_ϕ.

Remark 3.6: It should be noted that under Assumptions 3.1 **A1**), the space $\mathbb{F}$ of deterministic Markov decision rules is a compact set. Therefore, the search for optimal policies using ϕ-optimality (2) and (3) to be presented in the following may be refined by finding minimum rather than infimum of the set $\mathbb{F}$.

[3]A suitably adapted policy refers to an NVI adaptive policy to be introduced in Section IV.

Now we introduce the following assumption:

Assumption 3.2: For any $a = (\mathbf{r}, \mathbf{p}) \in \mathbb{A}$, $\mathbf{y}^f \in \{b_1, b_2\}^M$, $\mathbf{z} \in \mathbb{C}^M$, the matrix $\mathbf{\Psi}(\mathbf{y}^f, \mathbf{z}, \mathbf{r}, \mathbf{p}; \theta(\phi))\boldsymbol{\rho}'(\phi)$ is primitive and non-singular uniformly in ϕ.

Remark 3.7: Notice that Assumption 3.2 holds under very mild conditions for the noise process $\{\mathbf{w}_k\}_{k=1}^{\infty}$ when the transition probability matrix $\boldsymbol{\rho}(\phi)$ is primitive uniformly in ϕ. For further details see [20].

For each $\phi \in \Phi$, the ϕ-optimality equations (ϕ-OEs), also known as the Bellman equations,[4] associated with the adaptive average cost MDP problem $(\mathbf{P})$ may be expressed as

$$\min_{\lambda_\phi \in \mathbb{F}} \left[\sum_{\mathbf{y}^f, \acute{\mathbf{z}}} J_\phi^{\lambda_\phi}(\acute{s}) p(\acute{s}|s, a; \phi) - J_\phi^{\lambda_\phi}(s) \right] = 0 \tag{2}$$

$$J_\phi^{\lambda_\phi}(s) + v(s; \phi) = \min_{\lambda_\phi \in \mathbb{F}} \left[g(s, a; \phi) + \sum_{\mathbf{y}^f, \acute{\mathbf{z}}} v(\acute{s}; \phi) p(\acute{s}|s, a; \phi) \right] \tag{3}$$

where $a = \lambda_\phi(s) \in \mathbb{A}(s)$, $\acute{s} = (\acute{\alpha}(\mathbf{y}^f), \acute{\mathbf{z}})$, $J_\phi^{\lambda_\phi}$ is the average per-stage cost in steady state. Clearly, we are after its optimal value J_ϕ^*. $\sum_{\mathbf{y}^f, \acute{\mathbf{z}}} v(\acute{s}; \phi)\ p(\acute{s}|s, a; \phi)$ is called the cost-to-go function in which $v \in B(\mathbb{S} \times \Phi)$ with $B(\cdot)$ being the Banach space of real-valued bounded measurable functions v with the uniform norm $\| v \| \triangleq \sup_{s \in \mathbb{S}, \phi \in \Phi} | v(s; \phi) |$. The function $v(\cdot; \phi)$ is referred to as the differential cost defined as the expected total difference between per-stage cost $g(\cdot; \phi)$ and the stationary cost $J_\phi^{\lambda_\phi}$.

Remark 3.8: As mentioned at the end of Section II, in order to compute numerical solution to the ϕ-optimality equations, the forward variable $\acute{\alpha}$ is replaced by a discretized approximation which yields fully discretized ϕ-optimality equations. This, of course, depending on the choice of the discretization step leads to a suboptimal solution to the original problem. However, it can be shown following similar techniques used in [36] (see also [37]) that under a certain continuity condition on the differential cost associated with the Bellman equation, as the discretization step approaches zero, the optimal cost associated with the discretized ϕ-optimality equations converges to the optimal solution of the original continuous state space average cost optimality equation.

We also introduce the following additional assumptions from [28]:

Assumption 3.3: The cost-to-go function is a continuous function of $a = (\mathbf{r}, \mathbf{p}) \in \mathbb{A}(s)$ for every $s = (\alpha, \mathbf{z}) \in \mathbb{S}$, $\phi \in \Phi$, and $v \in B(\mathbb{S} \times \Phi)$.

Note that for the adaptive MDP $\mathcal{M}_\phi$ with a finite state space $\mathbb{S}$, Assumption 3.3 may be reduced to the following condition that the conditional probability measure $p(\acute{s}|s, a; \phi)$ is a continuous function of $a \in \mathbb{A}(s)$ for $\forall\, s, \acute{s} \in \mathbb{S}$, and $\forall \phi \in \Phi$.

In order to establish the existence of solutions to the ϕ-OEs for every $\phi \in \Phi$, an ergodicity condition was shown to hold in [28] (see Lemmas 3.1 and 3.2), which for a given $\phi \in \Phi$, under Assumption 3.2 implies that the average cost MDP problem $(\mathbf{P})$ associated with $\mathcal{M}_\phi$ forms a recurrent (ergodic) MDP. This means that the transition matrix corresponding to every deterministic stationary policy consists of a single recurrent class and no transient states. This is because using [28, Lemma 3.2], it can be shown[5] that the cost associated with every deterministic stationary Markov policy λ_ϕ is uniform in s, that is, $J_\phi^{\lambda_\phi}(s) = j_\phi^{\lambda_\phi} \in B(\Phi)$ for $\forall s \in \mathbb{S}$.

Thus we may characterize optimal policies and their corresponding average costs using only the single ϕ-optimality (3) as follows:

$$j_\phi^{\lambda_\phi} + v(s; \phi) = \min_{\lambda_\phi \in \mathbb{F}} \left[g(s, a; \phi) + \sum_{\mathbf{y}^f, \acute{\mathbf{z}}} v(\acute{s}; \phi) p(\acute{s}|s, a; \phi) \right] \tag{4}$$

where $a = \lambda_\phi(s) \in \mathbb{A}(s)$, and $\acute{s} = (\acute{\alpha}(\mathbf{y}^f), \acute{\mathbf{z}})$.

Now, we establish existence of solutions to the ϕ-optimality (4). Under Assumptions 3.1 and 3.3 and using the ergodicity results from [28], it can be shown from Corollary 3.6 in [22], that there exists a solution $\{j_\phi^*, v^*(\cdot; \phi)\}$ to the ϕ-optimality (4), where $j_\phi^* \in B(\Phi)$ is a real-valued bounded measurable function on Φ and $v^*(\cdot; \phi) : \mathbb{S} \times \Phi \mapsto \mathbb{R}$ is a real-valued bounded measurable function on $\mathbb{S}$ for each $\phi \in \Phi$. Based on this argument which insures the existence of a solution to the ϕ-OE, we present the following theorem:

Theorem 3.1: Suppose that Assumptions 3.1 and 3.3 hold, and there exist functions $j_\phi^* \in B(\Phi)$, and $v^*(\cdot; \phi) \in B(\mathbb{S} \times \Phi)$ in the Banach spaces of real-valued bounded measurable functions which satisfy the ϕ-OE (4). Assume there is a stationary deterministic Markov decision rule $\lambda_\phi^* \in \mathbb{F}$ which minimizes the right-hand side of the ϕ-OE (4), i.e., for each $\phi \in \Phi$ and $s \in \mathbb{S}$

$$j_\phi^* + v^*(s; \phi) = g(s, \lambda_\phi^*(s); \phi) + \sum_{\mathbf{y}^f, \acute{\mathbf{z}}} v^*(\acute{s}; \phi) p(\acute{s}|s, \lambda_\phi^*(s); \phi) \tag{5}$$

where $\lambda_\phi^*(s) \in \mathbb{A}(s)$. Then, the stationary policy λ_ϕ^* is average-optimal for the MDP $\mathcal{M}_\phi$, that is, action $a_k = (\mathbf{r}_k, \mathbf{p}_k) = \lambda_\phi^*(s_k)$ at time $k \geq 2$ determined by the stationary policy $\lambda_\phi^* = \{\lambda_\phi^*, \lambda_\phi^*, \ldots\}$ minimizes the cost J_ϕ defined in (1) and the value of the optimal cost is j_ϕ^*.

Remark 3.9: This theorem is essentially the ϕ-analog (parameterized version) of the existence theorem for optimal policies in average cost unichain models. See Proposition 1 in Chapter 8 in [38] or similarly Theorem 8.4.4 in [35].

D. Nonstationary Value Iteration

In this section, we develop the formulation for approximating MDP models $\mathcal{M}_{\hat{\phi}_k}$ and introduce a nonstationary value iteration (NVI) scheme and corresponding NVI adaptive Markov policies and show that under appropriate assumptions these adaptive policies are average optimal for the limit (true) MDP $\mathcal{M}_{\phi^\circ}$. The approach followed in this section is inspired by results on approximations and adaptive policies for average cost MDPs presented in [39] and further extended in [22].

Let $\{\hat{\phi}_k\}_{k=1}^{\infty}$ be a sequence in Φ converging to the true parameter ϕ° according to the following definition, where H_k refers to the vector space of admissible histories up to time k for $k \geq 0$, where $H_0 \triangleq \mathbb{S}$ and $H_k \triangleq \mathbb{K}^k \times \mathbb{S} = \mathbb{K} \times H_{k-1}$ for $k \geq 1$.

Definition 3.1: A sequence $\{\hat{\phi}_k\}_{k=1}^{\infty}$ of measurable functions $\hat{\phi}_k : H_k \mapsto \Phi$ is defined to be a sequence of strongly

[4] cf. [35] Sec. 8.4.

[5] cf. part (b) of Lemma 3.3 in [22].

consistent estimators of the true parameter ϕ° such that $\lim_{k\to\infty}\hat{\phi}_k = \phi^\circ \quad \mathbb{P}^{\lambda}_{s,\phi^\circ}$-a.s. is satisfied uniformly in λ for every $s \in \mathbb{S}$.

Remark 3.10: Note that there are several methods to estimate parameters of the HMM $\mathcal{H}$ in sense of the Definition 3.1. However, at this point, to maintain readability of the manuscript, it is simply assumed that the strongly consistent estimator $\{\hat{\phi}_k\}_{k=1}^{\infty}$ is available. This task is performed by the recursive ML parameter estimator module which is discussed in Section IV. Convergence of the ML estimator is then established in Section V where it is shown that the proposed recursive estimator is indeed strongly consistent.

Assumption 3.4: There are constants L_1 and L_2 such that the per-stage cost function $g(\phi)$ and the MDP transition kernel $p(\phi)$ satisfy the following inequalities uniformly in $\kappa = (s,a) \in \mathbb{K}$ for every $\phi, \acute{\phi} \in \Phi$:

$$|g(\kappa;\phi) - g(\kappa;\acute{\phi})| \le L_1 \bar{d}(\phi,\acute{\phi})$$
$$||p(\cdot|\kappa;\phi) - p(\cdot|\kappa;\acute{\phi})||_{tv} \le L_2 \bar{d}(\phi,\acute{\phi})$$

where $\bar{d}$ is the metric on the parameter space Φ.

In the following, we define adaptive policies as policies which determine the control actions adaptively based on the parameter estimates. First, we define nonstationary value iteration (NVI) functions $\bar{v}_k(s;\hat{\phi}_{k-1})$ recursively as follows.

Definition 3.2: Let $\bar{v}_0(\cdot)$ be an arbitrary function defined on $B(\mathbb{S}\times\Phi)$, e.g., $\bar{v}_0(\cdot) \triangleq 0$, and for every $s \in \mathbb{S}$, and $k \ge 1$

$$\bar{v}_1(s;\hat{\phi}_0) \triangleq \min_{a\in\mathbb{A}(s)} g(s,a;\hat{\phi}_0)$$
$$\bar{v}_{k+1}(s;\hat{\phi}_k) \triangleq \min_{a\in\mathbb{A}(s)} \Bigg[g(s,a;\hat{\phi}_k) + \sum_{\mathbf{y}^f,\acute{\mathbf{z}}} \bar{v}_k(\acute{s};\hat{\phi}_{k-1}) p(\acute{s}|s,a;\hat{\phi}_k) \Bigg] \quad (6)$$

where $\acute{s} = (\acute{\alpha}(\mathbf{y}^f), \acute{\mathbf{z}})$.

It is clear from Definition 3.2 that the NVI functions $\bar{v}_k(s;\hat{\phi}_{k-1})$ are obtained in an iterative manner starting from an arbitrary initial function $\bar{v}_0(\cdot)$ without the need to solve the Bellman ϕ-optimality (4) at each time step k, which in the case of our quantization problem is computationally intensive. This advantage makes the NVI scheme directly applicable to our problem. Note also that Assumption 3.4 is a sufficient condition for the NVI function $\bar{v}_k(s;\phi)$ is Lipschitz continuous in ϕ uniformly on $\mathbb{S}$ (see [28] for more details). The NVI adaptive policy corresponding to the NVI functions $\bar{v}_k(\cdot;\hat{\phi}_{k-1})$ is then defined as follows.

Definition 3.3: Let $\bar{\lambda} = \{\bar{\lambda}_k\}_{k=0}^{\infty}$, called the NVI adaptive policy, be a sequence of deterministic Markov decision rules, where for each $k \ge 0$, $\bar{\lambda}_k(\cdot;\hat{\phi}_k) \in \mathbb{F}$ is a measurable function such that action a determined by $a = \bar{\lambda}_k(s;\hat{\phi}_k) \in \mathbb{A}(s)$ minimizes the right-hand side of the ϕ-optimality (6) for every $s \in \mathbb{S}$. It is clear that the initial action at time $k = 0$ is determined by $\bar{\lambda}_0(s;\hat{\phi}_0) = \arg\min_{a\in\mathbb{A}(s)} g(s,a;\hat{\phi}_0)$.

The following theorem establishes the average optimality of the NVI adaptive policy $\bar{\lambda}$ for the true MDP $\mathcal{M}_{\phi^\circ}$.

Theorem 3.2: Suppose that Assumptions 3.1, 3.2, 3.3, and 3.4 hold. Let $\{\hat{\phi}_k\}_{k=0}^{\infty}$ be any sequence of measurable functions in Φ converging to the true parameter $\phi^\circ \quad \mathbb{P}^{\lambda}_{s,\phi^\circ}$-a.s. according to the Definition 3.1. Also, let $\bar{\lambda}_{\phi^\circ} = \{\bar{\lambda}_k\}_{k=0}^{\infty}$ be an adaptive policy as defined in Definition 3.3, where $\bar{\lambda}_k(s_k;\hat{\phi}_k(h_k)) \in \mathbb{F}$ for every $h_k \in H_k$. Then $\bar{\lambda}_{\phi^\circ}$ is an average-optimal policy for the true MDP $\mathcal{M}_{\phi^\circ}$.

Remark 3.11: Theorem 3.2 is the ϕ-analog of [22, Th. 6.6] for average cost MDPs. The proof follows from [22, Cor. 7.8, pp. 80] and the detail is omitted here.

IV. Recursive ML Parameter Estimator

A. Preliminaries

In this section, we develop the formulation for a recursive ML estimation of parameters of the HMM $\mathcal{H}$. The proposed recursive ML estimator is a measurable function in Φ which, at each time step k, finds the best estimate of the HMM parameters based on the MDP trajectory $h_k \in H_k$ until time k. The proposed method is an iterative (online[6]) variant of the generalized expectation maximization (EM[7]) algorithm for HMMs. The generalized EM algorithm uses the monotonicity property that the true likelihood increases at each iteration. Starting from some initial estimate, the EM algorithm iteratively finds the best estimate of the HMM parameters using ML criterion in two steps: an Expectation step (E-step) followed by a Maximization step (M-step). In the proposed algorithm, the E-step involves finding the distribution for the complete data given the known values for the observed (incomplete) data and estimates of the model parameters. The M-step finds new estimate of the parameters so as to increase the likelihood function (for further detail see [42] and [43]).

The proposed method is an adaptation of an online estimation algorithm based on relative entropy information measure presented in [10]. Hence, in the following we only present the variations necessary to our problem and all further details are omitted.

B. The Online EM Algorithm

In this section, we present an online EM algorithm which recursively estimates the parameters of the HMM $\mathcal{H} = (\boldsymbol{\rho}(\phi), \mathcal{S}(\phi), \pi, \boldsymbol{\Psi}(\theta(\phi)))$. Let $\hat{\phi}_k = (\boldsymbol{\rho}(\hat{\phi}_k), \mathcal{S}(\hat{\phi}_k), \hat{\theta}_k(\hat{\phi}_k)) \in \Phi$ be the estimate of model parameters at time $k \ge 0$. Let $\mathbf{O}_k^K \in \mathcal{O}_k^K$ denote the observable (incomplete) data at the fusion center from time instant k up to time K, where $\mathcal{O}_k^K$ is the σ-field generated by $(\mathbf{Y}_l^f, \mathbf{Z}_l, \mathbf{r}_l, \mathbf{p}_l)$, for $k \le l \le K$. For simplicity, we denote $\mathbf{O}_1^k$ by $\mathbf{O}^k$, and $\mathbf{O}_k^k$ by $\mathbf{O}_k$. Henceforth, for brevity, the state-to-observation probability distribution $\psi_i(\mathbf{y}_k^f, \mathbf{z}_k, \mathbf{r}_k, \mathbf{p}_k; \theta(\phi)),\ i = 1,\ldots,n$ may be denoted by $\psi_i(\mathbf{O}_k;\theta(\phi))$. Let $\varphi^k = \{\hat{\phi}_t\}_{t=0}^k$ denote the sequence of model estimates till time k based on the observations $\mathbf{O}^k$. Also, denote the sequence of unobservable (hidden) Markov chain states until time k by $X^k = \{X_t\}_{t=1}^k$. In the following, $l_c(\cdot)$ denotes a probability measure on $(\Omega,\mathcal{F})$ with respect to some σ-finite measure. It is shown in [10] that the M-step

[6]see [5], [40], [41] for the classic offline (nonrecursive) EM algorithm, known as Baum–Welch algorithm, for estimation of HMMs using the forward–backward procedure.

[7]see [42] for a general formulation of the EM algorithm and its basic properties.

of the online EM algorithm maximizes the relative entropy information measure which is equivalent to maximizing $\mathcal{J}(\phi) \triangleq \mathbb{E}_{\phi^\circ}\left[\log l_c(\mathbf{O}^k;\phi)\right]$, where $l_c(\mathbf{O}^k;\phi)$ is the marginal likelihood function of the observable (incomplete) data parameterized by ϕ. The M-step may be expressed as follows:

$$\begin{aligned}\hat{\phi}_k &= \arg\max_{\phi\in\Phi} \bar{Q}_k(\mathbf{O}^k, \varphi^{k-1};\phi), k\geq 1\\ \text{subject to}: &\quad \sum_{j=1}^{n} \varrho_{ij}(\phi) = 1, \quad \forall i = 1,\ldots,n\\ &\quad \varrho_{ij}(\phi)\geq 0, \quad \forall i,j = 1,\ldots,n \end{aligned} \tag{7}$$

where $\bar{Q}_k(\cdot)$ for $k\geq 1$ is computed in the E-step as follows:

E-step

$$\bar{Q}_k(\mathbf{O}^k,\varphi^{k-1};\phi) \triangleq \mathbb{E}_{X^k}\left[\log l_c(\mathbf{O}^k, X^k;\phi)|\mathbf{O}^k,\varphi^{k-1}\right]$$

where $l_c(\mathbf{O}^k, X^k;\phi)$ is the likelihood function of the complete data if X^k were fully observable.

Remark 4.1: Note that in the M-step optimization problem (7), depending on the distribution of the sensors' observations further constraints on the elements of the parameter vector ϕ might be required. As an example, for zero-mean Gaussian measurement noise processes $\{\mathbf{w}_k\}_{k=1}^{\infty}$, the standard deviation parameter $\theta^m(\phi)$ in the conditional density $f(\cdot|x_k;\theta^m(\phi))$ must be strictly positive. This can be ensured by introducing additional M number of constraints given by $\theta^m(\phi) > 0, \forall\, m = 1,\ldots,M$.

Define the constraint set $\mathbb{Z} \triangleq \{\phi : z_i(\phi)\leq 0,\ i = 1,\ldots,\ell\}$, where ℓ is the dimension of the parameter vector ϕ, and $z_i(\cdot) : \mathbb{R}^\ell \mapsto \mathbb{R}$ is a constraint defined as a real-valued C^1-function on $\mathbb{R}^\ell$. For our HMM parameter estimation problem, this constraint set includes the constraints on the transition probability elements as well as the noise density parameters as mentioned earlier. Without loss of generality, assume that $\nabla_\phi\, z_i(\phi)\neq 0$ for an active constraint $z_i(\cdot)$ at ϕ, that is, if $z_i(\phi) = 0$. Let $\mathcal{Z}(\acute{\phi})$ denote the set of indices of the active constraints at $\acute{\phi}$ defined by $\mathcal{Z}(\acute{\phi}) \triangleq \{i : z_i(\acute{\phi}) = 0\}$.

Remark 4.2: For tractability purposes, in our simulations, we shall deal with unconstrained optimization by considering a Lagrangian formulation. Thus, in order to insure positiveness of transition probabilities and standard deviation of the noise processes, we use a standard parameterization by considering square roots as $\phi = (\mathbf{S}(\phi), \mathcal{S}(\phi), \vartheta(\phi))$, where $\mathbf{S}(\phi) = [s_{ij}(\phi)]$ with $s_{ij} = \sqrt{\varrho_{ij}}$ for every $i, j = 1,\ldots,n$, and $\vartheta^m = \sqrt{\theta^m}$ for $m = 1,\ldots,M$. The optimization problem in M-step may then be expressed as follows:

$$\begin{aligned}\hat{\phi}_k = \arg\max_{\phi\in\Phi}\Bigg[&\bar{Q}_k(\mathbf{O}^k,\varphi^{k-1};\phi)\\ &-\sum_{i=1}^{n}\bar{\mu}_i\left(\sum_{j=1}^{n} s_{ij}^2(\phi) - 1\right)\Bigg],\\ k\geq 1&\end{aligned} \tag{8}$$

where the weighting factors $\bar{\mu}_i \geq 0$ are Lagrangian multipliers.

Definition 4.1: We introduce the following variables[8] which are used to implement the forward–backward procedure in order to evaluate $\bar{Q}_k(\cdot)$ in the EM algorithm:

$$\begin{aligned}\bar{\alpha}_k(i) &= \mathbb{P}_\phi(\mathbf{O}^k, X_k = \tilde{x}_i|\varphi^{k-1})\\ \bar{\beta}_{t|k}(i) &= \mathbb{P}_\phi(\mathbf{O}_{t+1}^k|X_k = \tilde{x}_i, \hat{\phi}_{t-1})\\ \bar{\gamma}_{t|k}(i) &= \mathbb{P}_\phi(X_t = \tilde{x}_i|\mathbf{O}^k,\varphi^{k-1}\\ \bar{\xi}_{t|k}(i,j) &= \mathbb{P}_\phi(X_t = \tilde{x}_i, X_{t+1} = \tilde{x}_j|\mathbf{O}^k,\varphi^{k-1})\end{aligned}$$

where $i,j = 1,\ldots,n$.

These variables can be recursively computed using standard recursions which can be found in [10], [13], and are not presented here due to space restrictions. By Theorem 3.2 suppose at each time step $k\geq 1$, action $(\mathbf{r}_k,\mathbf{p}_k)$ is determined according to a *deterministic* NVI adaptive policy $\bar{\lambda}_k(\cdot;\hat{\phi}_k)\in\mathbb{F}$. Further assume that the trajectory $\mathbf{O}^{k+1}\in\mathcal{O}^{k+1}$ has been observed. Then, for $k\geq 0$ the function $\bar{Q}_{k+1}(\cdot)$ may be evaluated as follows:

$$\begin{aligned}\bar{Q}_{k+1}(\mathbf{O}^{k+1},\varphi^k;\phi) = &\sum_{t=1}^{k+1}\chi_{t|k+1}(\phi)\\ &+\sum_{t=1}^{k+1}\bar{\chi}_{t|k+1}(\phi,\mathbf{O}_t)\\ &+\sum_{t=1}^{k}\log p_z(\mathbf{z}_{t+1}\mid\mathbf{z}_t)\\ &+\sum_{m=1}^{M}\sum_{i=1}^{u}\delta(z_{m,1}-\tilde{c}_i)\log\pi_i^m\end{aligned} \tag{9}$$

where $\delta(\cdot)$ is the Kronecker delta function, and the functions $\chi_{t|k+1}(\cdot)$ and $\bar{\chi}_{t|k+1}(\cdot)$ are evaluated as follows:

$$\begin{aligned}\chi_{t|k+1}(\phi) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\bar{\xi}_{t-1|k+1}(i,j)\log s_{ij}^2(\phi)\\ \bar{\chi}_{t|k+1}(\phi,\mathbf{O}_t) &= \sum_{i=1}^{n}\bar{\gamma}_{t|k+1}(i)\log\psi_i(\mathbf{O}_t;\vartheta(\phi)).\end{aligned} \tag{10}$$

Remark 4.3: From (9) we may write the following recursion[9] for the function $\bar{Q}_{k+1}(\cdot)$, $k\geq 1$:

$$\begin{aligned}&\bar{Q}_{k+1}(\mathbf{O}^{k+1},\varphi^k;\phi)\\ &\quad= \bar{Q}_k(\mathbf{O}^k,\varphi^{k-1};\phi) + \log p_z(\mathbf{z}_{k+1}\mid\mathbf{z}_k)\\ &\qquad+\chi_{k+1|k+1}(\phi)+\bar{\chi}_{k+1|k+1}(\phi,\mathbf{O}_{k+1}).\end{aligned} \tag{11}$$

We now present the following stochastic approximation algorithm which recursively adjusts the parameter vector ϕ by finding the (local) maximum of the objective function $\bar{Q}_k(\cdot)$ at each time step. Under appropriate regularity conditions introduced in Section V, the M-step of the online EM algorithm can be written as the following recursion:

$$M\text{-}\mathbf{step}: \hat{\phi}_{k+1} = \hat{\phi}_k + \epsilon_{k+1}(\phi)S(\hat{\phi}_k,\mathbf{O}_{k+1}),\ k\geq 0 \tag{12}$$

[8]Note that the (unnormalized) forward variable $\bar{\alpha}_k$ has been introduced earlier in Lemma 2.1.

[9]In the following, this recursion is used in evaluating the score function.

where $\epsilon_{k+1}(\phi)$ is a sequence of decreasing small scalar gains which satisfy certain well-known conditions[10] as follows:

$$\epsilon_k(\phi) \geq 0, \quad \sum_k \epsilon_k(\phi) = +\infty, \quad \sum_k \epsilon_k^2(\phi) < \infty \tag{13}$$

and $S(\hat{\phi}_k, \mathbf{O}_{k+1})$ is the score function defined as

$$S(\hat{\phi}_k, \mathbf{O}_{k+1}) \triangleq \nabla_\phi \, \bar{Q}_{k+1}(\mathbf{O}^{k+1}, \varphi^k; \phi)\Big|_{\phi=\hat{\phi}_k}. \tag{14}$$

Remark 4.4: Using recursion (11), it is straightforward to show that the incremental score function can be computed as follows:

$$S(\hat{\phi}_k, \mathbf{O}_{k+1}) = \nabla_\phi \big(\chi_{k+1|k+1}(\phi) + \bar{\chi}_{k+1|k+1}(\phi, \mathbf{O}_{k+1})\big)\Big|_{\phi=\hat{\phi}_k}. \tag{15}$$

A detailed summary of the estimation algorithms for the various HMM parameters are provided in the Appendix.

Remark 4.5: Note that $\epsilon_k(\phi)$ is chosen to be a function of the parameter type, that is, Markov chain transition probabilities, state levels, and measurement noise of the sensors each has different gain values. This improves the convergence rate of the parameters. Convergence analysis of the above algorithm given by (12), (15) can be carried out using the standard mean ODE approach as in [44] under some regularity assumptions on the observation probability distributions along with an additional geometric ergodicity condition on an extended Markov chain [16]. A detailed analysis is presented in Section V.

V. Asymptotic Analysis of the On-Line ML Estimator

In this section, we analyze the asymptotic behavior of the parameter estimation algorithm presented in the previous section. In particular, we study the convergence of the estimation algorithm (12) using the ordinary differential equation (ODE) method [27], [44]. The idea behind the ODE approach is to show that, under appropriate conditions on the noise and using suitably adopted step-size ϵ_k, the asymptotic behavior of the difference (12) can be efficiently determined by analyzing stability of a mean ODE. In this way, the asymptotics of the iterate sequence $\{\hat{\phi}_k\}$ can be studied by analyzing the limit trajectory or equilibrium point(s) of a continuous time process found by interpolating the iterates $\{\hat{\phi}_k\}$ with interpolation intervals $\{\epsilon_k\}$ [27]. In the following, we follow a related approach in [16] by showing that their sufficient conditions hold in case of our algorithm. In order to maintain rigor and completeness, we explicitly state the set of assumptions and the necessary intermediate results in the context of our recursive parameter estimation algorithm.

A. Preliminary Assumptions

Define the vector $\psi(\mathbf{O}_k, \phi)$ and the matrix $\mathbf{\Psi}(\mathbf{O}_k, \phi)$ by

$$\psi(\mathbf{O}_k, \phi) \triangleq [\psi_1(\mathbf{O}_k; \theta(\phi)), \ldots, \psi_n(\mathbf{O}_k; \theta(\phi))]'$$
$$\mathbf{\Psi}(\mathbf{O}_k, \phi) \triangleq \mathrm{diag}[\psi_1(\mathbf{O}_k; \theta(\phi)), \ldots, \psi_n(\mathbf{O}_k; \theta(\phi))].$$

We introduce the following assumptions:

Assumption 5.1: For any $k \geq 1$, and any $\tilde{x}_i \in \mathcal{S}$, and for a given $(\mathbf{r}, \mathbf{p}) \in \mathbb{A}$, and $\mathbf{z} \in \mathbb{C}^M$, the conditional probability distribution of the observable data $\mathbf{y}_k^f$ given the realization $X_k = \tilde{x}_i$ is absolutely continuous with respect to a nonnegative and σ-finite measure μ, with μ-a.e. positive densities $\psi_i(\mathbf{y}_k^f, \mathbf{z}, \mathbf{r}, \mathbf{p}; \theta(\phi))$ uniformly in ϕ.

Remark 5.1: In case of our quantization problem, μ is a counting measure on $\mathbb{Y} \triangleq \{b_1, b_2\}^M$. Assumption 5.1 is satisfied if the probability of receiving a particular $\mathbf{Y}_{m,k}^f \in (b_1, b_2)$ (from the mth sensor) at the fusion center given the Markov chain is in state i, is positive. Note again that this is satisfied for i.i.d. Gaussian sensor measurement noise and the assumed ergodic properties of the independent finite-state Markov communication channels between the sensors and the fusion center.

Remark 5.2: Under Assumptions 3.2 and 5.1, the Markov process $\{X_k, \mathbf{O}_k\}_{k=1}^\infty$ is geometrically ergodic on the state space $\mathcal{S} \times \mathbb{O}$ under the measure $\mathbb{P}_{\phi^\circ}$, where $\mathbb{O}$ is defined by $\mathbb{O} \triangleq \mathbb{Y} \times \mathbb{C}^M \times \mathbb{A}$, see [45].

Assumption 5.2: $\boldsymbol{\rho}(\phi)$ is a real-valued C^2-function with bounded first and second derivatives denoted by $\boldsymbol{\rho}^{(1)}(\phi)$ and $\boldsymbol{\rho}^{(2)}(\phi)$, respectively, satisfying the following inequality:

$$|\boldsymbol{\rho}^{(2)}(\phi) - \boldsymbol{\rho}^{(2)}(\acute{\phi})| \leq L_3 \bar{d}(\phi, \acute{\phi}) \tag{16}$$

where $L_3 \geq 0$ is a constant, and $\bar{d}$ denotes a suitable metric on the parameter space Φ.

Assumption 5.3: For any $\mathbf{o} \in \mathbb{O}$, $\psi(\mathbf{o}, \cdot)$ is a real-valued C^3-function on Φ.

Remark 5.3: Assumption 5.3 is satisfied if the marginal densities of the innovations $\{\mathbf{w}_k\}_{k=1}^\infty$ are continuous with their derivatives being bounded with respect to ϕ. Clearly, this holds for innovations with normal distributions.

For every $k \geq 2$, let $\nu_k(\phi) = [\nu_k^i(\phi)]$ denote the n-dimensional prediction filter vector of the Markov process state $\{X_k\}_{k=1}^\infty$, with its ith element being defined by

$$\nu_k^i(\phi) \triangleq \mathbb{P}_\phi(X_k = \tilde{x}_i | \mathbf{O}^{k-1}), \quad i = 1, \ldots, n$$

where $\nu_k^i(\phi)$ denotes the conditional density of the state X_k at time k given a realization of the observed data until time $k-1$. It can be shown [46] that the prediction filter satisfies the following recursion known as the forward Baum equation

$$\nu_{k+1}(\phi) = \frac{\boldsymbol{\rho}'(\phi) \mathbf{\Psi}(\mathbf{O}_k, \phi) \nu_k(\phi)}{\psi'(\mathbf{O}_k, \phi) \nu_k(\phi)}, \quad k \geq 1$$

with the induction being initialized by $\nu_1 = \pi$. The random sequence $\{\nu_k\}_{k=1}^\infty$ takes values in the simplex $\mathcal{P}(\mathcal{S})$, defined as the set of probability distributions over the state space $\mathcal{S}$ with respect to the Lebesgue measure $\mho$.

Define the random sequence $\{\omega_k(\phi)\}_{k=1}^\infty$, where $\omega_k(\phi) \triangleq [\omega_k^1(\phi), \ldots, \omega_k^\ell(\phi)]$ is the $n \times \ell$ matrix of partial derivatives of $\nu_k(\phi)$ with respect to components of the parameter vector ϕ, with its ith column defined by $\omega_k^i(\phi) \triangleq \nabla_{\phi_i} \nu_k(\phi)$, $1 \leq i \leq \ell$. It is clear that $\omega_k(\phi)$ resides in a set Ξ defined as $\Xi \triangleq \{\omega \in \mathbb{R}^{n \times \ell} : \mathbb{1}_n' \omega = \mathbb{0}_\ell'\}$, where $\mathbb{0}_\ell$ denotes the ℓ-dimensional column vector with all elements equal to zero.

Let $\bar{\nu}_k(\phi) = [\bar{\nu}_k^i(\phi)]$ denote the n-dimensional filter vector of the Markov process state $\{X_k\}_{k=1}^\infty$, with its ith element defined by $\bar{\nu}_k^i(\phi) \triangleq \mathbb{P}_\phi(X_k = \tilde{x}_i | \mathbf{O}^k)$. Similar to the prediction filter, define $\bar{\omega}_k(\phi) \triangleq [\bar{\omega}_k^1(\phi), \ldots, \bar{\omega}_k^\ell(\phi)]$ as the $n \times \ell$ matrix of partial derivatives of $\bar{\nu}_k(\phi)$ with respect to components of the parameter vector ϕ.

[10] see [44] for further details.

Assumption 5.4: Under the probability measure $\mathbb{P}_{\phi^\circ}$, the extended Markov process $\{X_k, \mathbf{O}_k, \bar{\nu}_k(\phi), \bar{\omega}_k(\phi)\}_{k=1}^{\infty}$ is geometrically ergodic on the state space $\mathcal{S}\times\mathbb{O}\times\mathcal{P}(\mathcal{S})\times\Xi$. As a result, it has a unique invariant probability measure $\bar{\eta}_\phi$ under the measure $\mathbb{P}_{\phi^\circ}$.

Remark 5.4: Under Assumption 5.4, the initial conditions $\bar{\nu}_1$ and $\bar{\omega}_1$ are forgotten exponentially fast and as such they are asymptotically trivial in our analysis.

Remark 5.5: Geometric ergodicity of the Markov process $\{X_k, \mathbf{O}_k, \nu_k(\phi), \omega_k(\phi)\}_{k=1}^{\infty}$, where $\nu_k(\phi), \omega_k(\phi)$ corresponds to the prediction filter, has been studied in [45]. We briefly mention their sufficient conditions in the following. Define for any $e \geq 0$

$$\varepsilon^{(0)}(\mathbf{o}) \triangleq \sup_{\phi\in\Phi} \frac{\max_{i\in\{1,\ldots,n\}}\psi_i(\mathbf{o};\theta(\phi))}{\min_{i\in\{1,\ldots,n\}}\psi_i(\mathbf{o};\theta(\phi))}$$

$$\Delta_e^{(0)} \triangleq \sup_{\phi\in\Phi} \max_{i\in\{1,\ldots,n\}} \sum_{\mathbf{o}\in\mathbb{O}} [\varepsilon^{(0)}(\mathbf{o})]^e \psi_i(\mathbf{o};\theta(\phi)).$$

Then, the sufficient condition is that Assumptions 3.2 and 5.1 hold and $\Delta_2^{(0)}$ and $\Delta_4^{(0)}$ are both finite. It can be shown that Assumption 5.4 holds under the same sufficient conditions, see also [47]. Note that for any $e \geq 0$, $\Delta_e^{(0)}$ is finite when the innovations $\{\mathbf{w}_k\}_{k=1}^{\infty}$ are i.i.d. Gaussian random processes. Note that the i.i.d. Gaussian measurement noise and the ergodicity of the Markov channel process ensure the continuity and positivity assumption needed to make the transition probabilities of state to (quantized) measurements received at the fusion center all strictly positive. It is then easy to see that if in addition, the Markov chain transition probability matrix $\boldsymbol{\rho}(\phi)$ is primitive for all ϕ, Assumption 5.4 will be satisfied.

B. Kullback–Leibler Measure

Define the $n \times n$ matrix $\mathbf{L}(\phi) \triangleq [\varrho_{ij}(\phi)\log\varrho_{ij}(\phi)]$. From recursion (11), at each time step $k \geq 1$, the function $\bar{Q}_k(\cdot)$ (suitably normalized) in the online EM estimation algorithm (7), may be expressed as

$$\bar{Q}_k(\phi) = \frac{1}{k}\sum_{t=1}^{k} \chi_{t|t}(\phi) + \bar{\chi}_{t|t}(\phi, \mathbf{O}_t), \quad k \geq 1$$

or equivalently in terms of the filter vector $\bar{\nu}_k(\phi)$

$$\bar{Q}_k(\phi) = \frac{1}{k}\sum_{t=1}^{k}\Big[\langle \boldsymbol{\Psi}(\mathbf{O}_t,\phi)\mathbf{L}'(\phi)\bar{\nu}_{t-1}(\phi),\ \mathbb{1}_n\rangle + \log[\psi'(\mathbf{O}_t,\phi)]\bar{\nu}_t(\phi)\Big]. \quad (17)$$

Define for any $\imath \geq 0$

$$\bar{\Gamma}_\imath \triangleq \sup_{\phi\in\Phi}\max_{l\in\{1,\ldots,n\}}\sum_{\mathbf{o}\in\mathbb{O}}\left[\max_{i,j\in\{1,\ldots,n\}}|\varrho_{ij}(\phi)\psi_j(\mathbf{o};\theta(\phi))\log\varrho_{ij}(\phi)|\right]^\imath \times \psi_l(\mathbf{o};\theta(\phi^\circ))$$

$$\tilde{\Gamma}_\imath \triangleq \sup_{\phi\in\Phi}\max_{i\in\{1,\ldots,n\}}\sum_{\mathbf{o}\in\mathbb{O}}\left[\max_{j\in\{1,\ldots,n\}}|\log\psi_j(\mathbf{o};\theta(\phi))|\right]^\imath \psi_i(\mathbf{o};\theta(\phi^\circ)). \quad (18)$$

It can be shown that if Assumptions 3.2, 5.1, and 5.4 hold and $\Delta_1^{(0)}$, $\bar{\Gamma}_1$, and $\tilde{\Gamma}_1$ are finite, then the functions $\chi_{t|t}(\phi)$ and $\bar{\chi}_{t|t}(\phi, \mathbf{O}_t)$ are locally Lipschitz continuous uniformly on $\mathbb{O}$. See [45] for a proof in a similar case. Using this and Assumptions 3.2, and 5.1–5.4, suppose $\Delta_1^{(0)}$, $\bar{\Gamma}_1$, and $\tilde{\Gamma}_1$ are finite, then $\lim_{k\to\infty}\bar{Q}_k(\phi) = \bar{Q}(\phi)$ $\mathbb{P}_{\phi^\circ}$-a.s. is satisfied uniformly in ϕ, where

$$\bar{Q}(\phi) = \sum_{\mathbf{o}\in\mathbb{O}}\int_{\mathcal{P}(\mathcal{S})}\Big[\langle\boldsymbol{\Psi}(\mathbf{o},\phi)\mathbf{L}'(\phi)\bar{\nu}(\phi),\ \mathbb{1}_n\rangle + \log[\psi'(\mathbf{o},\phi)]\bar{\nu}(\phi)\Big] \times \bar{\eta}_\phi^{\mathcal{P}\mathbb{O}}(\mathbf{o}, d\bar{\nu})$$

in which $\bar{\eta}_\phi^{\mathcal{P}\mathbb{O}}$ denotes the marginal density function of the invariant measure $\bar{\eta}_\phi$ defined on $\mathbb{O}\times\mathcal{P}(\mathcal{S})$.

It has been shown [48] that the true parameter ϕ° is an element of $\mathbb{L}_{KL} \triangleq \arg\min_{\phi\in\Phi}\mathbf{K}(\phi)$ defined as the set of global minima of the Kullback–Leibler information measure $\mathbf{K}(\phi)$ defined by $\mathbf{K}(\phi) \triangleq -[\bar{Q}(\phi) - \bar{Q}(\phi^\circ)] \geq 0$.

From (14) and (17), the incremental score function $S(\phi)$ can be expressed in terms of the extended observation $\mathbf{W}_k \triangleq (\mathbf{O}_k, \bar{\nu}_k(\phi), \bar{\omega}_k(\phi))$ which includes the filter vector and its derivative as follows:

$$S(\phi,\mathbf{W}_k) = \nabla_\phi\Big[\langle\boldsymbol{\Psi}(\mathbf{O}_k,\phi)\mathbf{L}'(\phi)\bar{\nu}_{k-1}(\phi),\ \mathbb{1}_n\rangle + \log[\psi'(\mathbf{O}_k,\phi)]\bar{\nu}_k(\phi)\Big]. \quad (19)$$

C. Convergence Analysis

In this section, we provide convergence analysis of the estimation algorithm given in (12). Using (19) and considering all the required constraints on the parameter ϕ, we may rewrite the estimation algorithm given in (12) in the form

$$\hat{\phi}_{k+1} = \Pi_{\mathbb{Z}}[\hat{\phi}_k + \epsilon_{k+1}(\phi)S(\hat{\phi}_k, \mathbf{W}_{k+1})], \quad k \geq 0 \quad (20)$$

where $\mathbb{Z}$, called the constraint set defined earlier, is assumed to be a nonempty, convex and compact set, and $\Pi_{\mathbb{Z}}$ is the projection of the parameter estimate to the constraint set $\mathbb{Z}$. The recursion (20) can be written as the following stochastic approximation algorithm:

$$\hat{\phi}_{k+1} = \hat{\phi}_k + \epsilon_{k+1}(\phi)S(\hat{\phi}_k, \mathbf{W}_{k+1}) + \epsilon_{k+1}(\phi)V_{k+1} \quad (21)$$

where V_{k+1} defines a small residual perturbation (also known as the correction or projection term) on the algorithm necessary to confine the parameter estimate to the constraint set $\mathbb{Z}$ if it ever slips away from $\mathbb{Z}$, see [27, p. 121].

In order to analyze the asymptotics of the iterate sequence $\{\hat{\phi}_k\}$ in (20), we define a projected ODE for a continuous time process $\{\phi(t); t \geq 0\}$ obtained by interpolating the iterates $\{\hat{\phi}_k\}$ with interpolation intervals $\{\epsilon_k\}$ as follows:

$$\dot{\phi} = \mathbf{H}(\phi) + \tilde{z}, \quad \phi(0) = \hat{\phi}_0,\ \tilde{z} \in -\mathcal{C}(\phi) \quad (22)$$

where $\tilde{z}(\cdot)$ is the projection or correction term defined as the minimum distance needed to bring $\phi(t)$ to the constraint set $\mathbb{Z}$, and $\mathcal{C}(\phi)$ is the convex cone generated by the set of outward normals $\{\nabla_\phi z_i(\phi) : i \in \mathcal{Z}(\phi)\}$. The function $\mathbf{H}(\phi)$, known as mean vector field, is defined by $\mathbf{H}(\phi) \triangleq \lim_{k\to\infty}\mathbb{E}_\phi[S(\phi,\mathbf{W}_k)] = \nabla_\phi\mathbf{K}(\phi)$. For further details on projected ODE, see [27].

Assumption 5.5: For each $\phi \in \mathbb{Z}$, $S(\phi, \mathbf{W}_k)$ is uniformly integrable, the function $\mathbf{H}(\phi)$ is a C^0-function in ϕ, and also for each $\mathbf{W}$, $S(\cdot, \mathbf{W})$ is a C^0-function on $\mathbb{Z}$.

Assumption 5.6: There exists nonnegative measurable functions $L(\cdot)$ and $D(\cdot)$, where $D(\cdot)$ is bounded on bounded ϕ-set, such that

$$\mid S(\phi, \mathbf{W}) - S(\acute{\phi}, \mathbf{W}) \mid \leq L(\mathbf{W}) D(\phi, \acute{\phi})$$

and $D(\phi) \to 0$ as $\phi \to 0$ and such that for some integer $\tilde{t} \geq 0$

$$\mathbb{P}\left(\limsup_{t \geq 1} \sum_{k=t}^{t+\tilde{t}} \epsilon_k L(\mathbf{W}_k \right) < \infty) = 1.$$

Define for any $\imath \geq 0$

$$\varepsilon^{(1)}(\mathbf{o}) \triangleq \sup_{\phi \in \Phi} \max_{l \in \{1,\ldots,\ell\}} \frac{\max_{i \in \{1,\ldots,n\}} \mid \nabla_{\phi l} \psi_i(\mathbf{o}; \theta(\phi)) \mid}{\min_{i \in \{1,\ldots,n\}} \psi_i(\mathbf{o}; \theta(\phi))}$$

$$\Delta_\imath^{(1)} \triangleq \sup_{\phi \in \Phi} \max_{i \in \{1,\ldots,n\}} \sum_{\mathbf{o} \in \mathbb{O}} [\varepsilon^{(1)}(\mathbf{o})]^\imath \psi_i(\mathbf{o}; \theta(\phi))$$

$$\Upsilon_\imath \triangleq \sup_{\phi \in \Phi} \max_{i \in \{1,\ldots,n\}} \sum_{\mathbf{o} \in \mathbb{O}} \mid \mathbf{o} \mid^\imath \psi_i(\mathbf{o}; \theta(\phi)).$$

Remark 5.6: A sufficient condition for Assumptions 5.5 and 5.6 is that $\Delta_2^{(1)}$, Υ_2, $\bar{\Gamma}_2$, and $\tilde{\Gamma}_2$ are finite. It is clear that this is satisfied for the innovations $\{\mathbf{w}_k\}$ which are i.i.d. Gaussian random processes, see [15]. Essentially, Assumptions 5.5 and 5.6 impose uniform integrability and Lipschitz continuity on the incremental score function $S(\phi, \mathbf{W})$ [see (19)], given by the derivative of the per stage cost in the normalized expected log-likelihood function $\bar{Q}_k(\phi)$, which is maximized via the recursive ML algorithm to find the parameter estimates.

Note that according to Lyapunov stability, a set $\bar{\mathbb{Z}} \subset \mathbb{Z}$ is locally asymptotically stable for the ODE (22) if for each $\iota > 0$ there is a $\iota_1 > 0$ such that all trajectories $\phi(t)$ starting in $\mathbb{B}_{\iota_1}(\bar{\mathbb{Z}})$ never leave $\mathbb{B}_\iota(\bar{\mathbb{Z}})$ and ultimately go to $\bar{\mathbb{Z}}$, where $\mathbb{B}_\iota(\bar{\mathbb{Z}})$ denotes an ι neighborhood of $\bar{\mathbb{Z}}$. We now make the following assumption:

Assumption 5.7: Suppose that the set $\mathbb{L}_{KL}$ is locally asymptotically stable for the ODE (22). For any initial condition $\hat{\phi}_0 \notin L_{\mathbb{Z}}^1$, for $L_{\mathbb{Z}}^1 \subseteq L_{\mathbb{Z}}$, where $L_{\mathbb{Z}}$ is the set of limit point(s) of the mean ODE (22), the trajectories of (22) goes to $\mathbb{L}_{KL}$.

Theorem 5.1: Suppose Assumptions 3.2, and 5.1–5.7 hold. Then, $\hat{\phi}_k$ converges to $\phi^\circ \quad \mathbb{P}_{\phi^\circ}$-a.s. as $k \to \infty$.

Remark 5.7: The proof follows from [16, Th. 3.4], see also [27, Ch. 6, Th. 1.1].

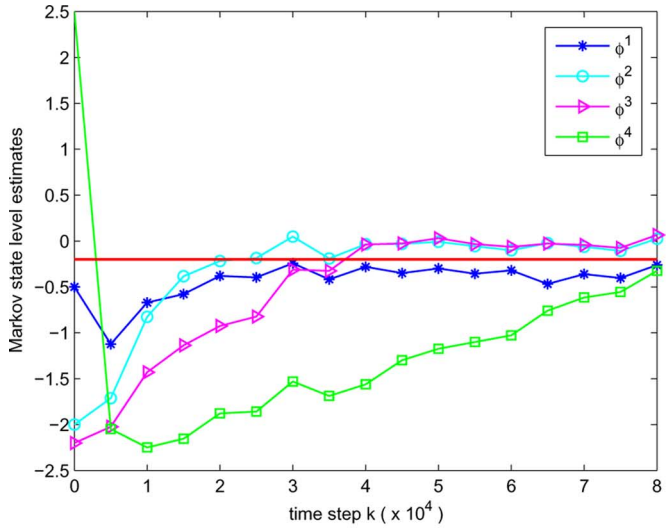
VI. Performance Evaluation

In this section, numerical results are presented to illustrate the performance of the proposed joint optimal quantization and power allocation and HMM estimation algorithm. We study the effect of initial estimates on the convergence of HMM parameters and optimal cost, and illustrate the effect of distance and channel quality on the performance of the ML estimator. Also, simulations have been performed for different values of the tradeoff parameter β (corresponding to various values of average sum powers) to illustrate the performance of our estimation scheme compared to the case when we have exact knowledge of the true parameters. Unless otherwise mentioned, the variables which are assumed fixed throughout the following experiments are as follows: the step size in discretizing the information state $\tilde{\alpha}_{k+1;\hat{\phi}(k)}$ is 0.01, the path loss exponent of the wireless channel is considered $\varsigma = 2$ (for example, in an indoor factory environment [49]), and the constant coefficient γ for computing crossover probabilities is $\gamma = 2$.

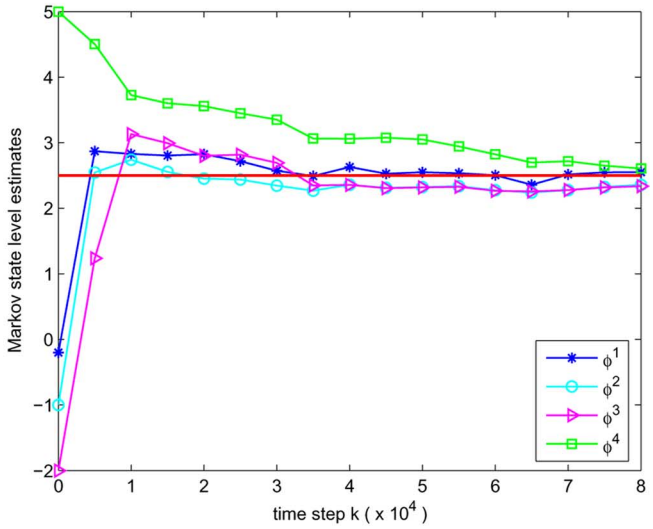
First we illustrate the performance of the online EM parameter estimation algorithm. For these simulations, we generated random sequences of 80 000 observations obtained by two sensors measuring a two-state Markov chain $\{X_k\}_{k=1}^\infty$ with state space $\mathcal{S}(\phi^\circ) = \{-0.2, 2.5\}$ and transition kernel $\boldsymbol{\rho}(\phi^\circ) = \begin{bmatrix} 0.94 & 0.06 \\ 0.28 & 0.72 \end{bmatrix}$. The measurement noises $\{w_{m,k}\}_{k=1}^\infty$ of the sensors are assumed to be zero-mean white Gaussian noise processes with a noise variance vector $\sigma^2(\phi^\circ) = [0.5, 0.3]'$. The sensors are located at different distances from the fusion center with distance vector $d = [80.0, 180.0]'$, where the figures are given in meters. Therefore, in the following experiments, the parameter vector based on usual parameterization is defined as $\phi = (\tilde{x}_1, \tilde{x}_2, \varrho_{11}, \varrho_{12}, \varrho_{21}, \varrho_{22}, \theta^1, \theta^2)'$ and the true parameter is given by $\phi^\circ = (-0.2, 2.5, 0.94, 0.06, 0.28, 0.72, \sqrt{0.5}, \sqrt{0.3})'$. Note that, however, as mentioned in Remark 4.2 in order to deal with constraints the actual parameters to be estimated are based on new parameterization as $\phi = (\tilde{x}_1, \tilde{x}_2, s_{11}, s_{12}, s_{21}, s_{22}, \vartheta^1, \vartheta^2)'$. The wireless channels from the sensors to the fusion center are assumed to be independent and each channel is modeled by a two state Markov chain with state space $\mathbb{C} = \{\tilde{c}_1, \tilde{c}_2\}$. The channel states $\tilde{c}_1$ and $\tilde{c}_2$ represent the corresponding channel gains $g_1^2 = 3 \times 10^{-8}$ and $g_2^2 = 2 \times 10^{-9}$, respectively, where clearly g_2 corresponds to a "worse" channel state. The channels are assumed to be asymmetric across the two sensors, that is, having different fading statistics with the transition probability matrices given by $\mathbf{C}^1 = \begin{bmatrix} 0.66 & 0.34 \\ 0.61 & 0.39 \end{bmatrix}$, $\mathbf{C}^2 = \begin{bmatrix} 0.79 & 0.21 \\ 0.32 & 0.68 \end{bmatrix}$. The noise power of the wireless channel for each sensor is $\sigma_v^2 = 3 \times 10^{-14}\,\mathrm{W}$. The power levels for each sensor is chosen from the action space $\mathbb{V} = \{65, 30, 10\}$, with the figures being in mW. The action space of the quantization thresholds for each sensor is given by the finite set $\mathbb{U} = \{-0.8, 0.7, 1.8, 3.5\}$. Note that the action space for the sensor transmission powers and quantization threshold levels can actually be (locally) optimized by using a gradient-free stochastic optimization approach [based on an adaptive simultaneous perturbation based stochastic approximation (SPSA)], as was done by us in a previous paper [20]. However, we have not explored this aspect in the current paper in order to keep the computational complexity low. The tradeoff parameter for these initial experiments is set to $\beta = 0$ (thus corresponding to no constraint on the average sum power).

As mentioned earlier, the gain sequence $\{\epsilon_k(\phi)\}_{k=1}^\infty$ must satisfy certain sufficient conditions given in (13) in order to ensure almost sure convergence of the model parameters. These conditions essentially mean that ϵ_k should tend to zero as $k \to \infty$ at a rate neither too fast nor too slow. We picked $\epsilon_k(\phi) = 1/(\upsilon_1 + k)^{\upsilon_2(\phi)}$, where υ_2 takes different values[11] depending on the parameter type. The typical values chosen in our numerical examples are $\upsilon_1 = 50$, $\upsilon_2(\mathbf{S}) = 0.7$, $\upsilon_2(\mathcal{S}) = 0.25$, $\upsilon_2(\vartheta) = 0.62$.

[11] υ_1 may also be chosen based on the parameter type to further improve the convergence rate.



(a)



(b)

Fig. 4. Convergence of the state level parameters (a) $\bar{x}_1(\hat{\phi}_k)$ and (b) $\bar{x}_2(\hat{\phi}_k)$ for various initial conditions. The true values $\bar{x}_1(\phi^\circ) = -0.2$ and $\bar{x}_2(\phi^\circ) = 2.5$ are marked by the red lines.

As for other implementation aspects, we have used fixed-lag estimation schemes to approximate the fixed interval variables $\bar{\beta}_{t|k+1}, \bar{\gamma}_{t|k+1}, t \leq k+1$, and $\bar{\xi}_{t|k+1}, t \leq k$, by fixed-lag variables[12] $\bar{\beta}_{t|t+\tau}, \bar{\gamma}_{t|t+\tau}$, and $\bar{\xi}_{t|t+\tau}$ with sufficiently large τ chosen to be $\tau = 200$. This is done to reduce computational complexity of the online algorithm on a finite but very long observation sequence.

For the Markov chain state levels $\tilde{x}_1(\phi)$ and $\tilde{x}_2(\phi)$ we have examined four different initial estimates in the range up to $\pm 3\sigma$, where $\sigma \triangleq \max_m\{\sigma_m(\phi^\circ)\}$, away from the true values $\tilde{x}_1(\phi^\circ)$, and $\tilde{x}_2(\phi^\circ)$. The initial estimates for the state level parameters according to usual parameterization are given as

$$\phi^1 = (-0.5, -0.2, 0.75, 0.25, 0.4, 0.6, 6.25, 0.25)'$$
$$\phi^2 = (-2.0, -1.0, 0.75, 0.25, 0.4, 0.6, 6.25, 0.25)'$$
$$\phi^3 = (-2.2, -2.0, 0.75, 0.25, 0.4, 0.6, 6.25, 0.25)'$$
$$\phi^4 = (2.5, 5.0, 0.75, 0.25, 0.4, 0.6, 6.25, 0.25)'.$$

[12]For further details on how to evaluate the fixed-lag variables refer to [10] Section IV.

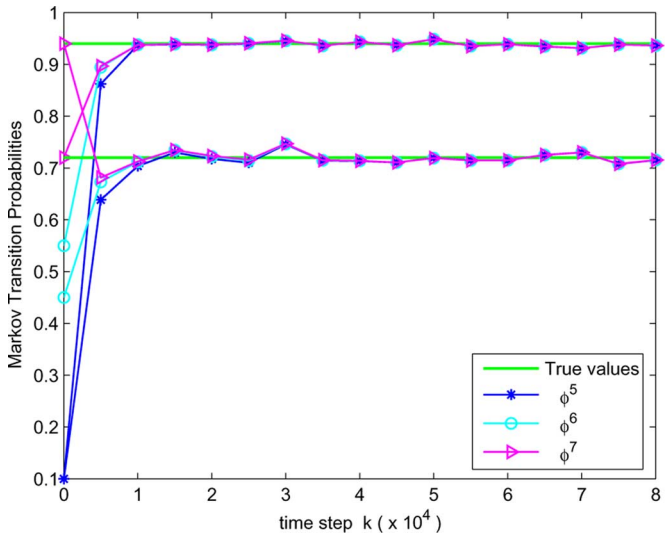


Fig. 5. Convergence of the transition probabilities $\varrho_{11}(\hat{\phi}_k)$ and $\varrho_{22}(\hat{\phi}_k)$ for different initial estimates. The green lines represent the true values.

Fig. 4 shows the effect of the above initial estimates on the convergence of state level parameters $\tilde{x}_1(\hat{\phi}_k)$ and $\tilde{x}_2(\hat{\phi}_k)$, respectively. The estimates are averaged over each 5×10^3 time steps. Convergence is achieved in all cases, although it is slower in cases where more than one initial estimates are $\pm 3\sigma$ or further apart from the true values as, for example, in ϕ^4. Even in cases where the initial condition for one state level is close or equal to a different true state level, we still achieve a relatively fast convergence.

Fig. 5 shows the convergence of the transition probabilities $\varrho_{11}(\hat{\phi}_k)$ and $\varrho_{22}(\hat{\phi}_k)$ under several initial conditions. The initial estimates are

$$\phi^5 = (-1, 1.5, 0.10, 0.90, 0.90, 0.10, 6.25, 0.25)'$$
$$\phi^6 = (-1, 1.5, 0.55, 0.45, 0.55, 0.45, 6.25, 0.25)'$$
$$\phi^7 = (-1, 1.5, 0.72, 0.28, 0.06, 0.94, 6.25, 0.25)'.$$

For all these initial conditions, fast convergence have been achieved even for ϕ^7 where more than one initial estimate is equal to the true parameter value. Note that the initial point ϕ^5 is chosen such that both ϱ_{11} and ϱ_{22} start from the same point 0.1 which is considerably away from their true values.

Fig. 6 shows the convergence of the NVI adaptive cost $j_{\hat{\phi}_k}^{\bar{\lambda}_k}$ for various initial conditions. It is clear that ϕ^4 has slower convergence regarding NVI cost because convergence of its parameters is also achieved in a slower rate as can be seen in Fig. 4. Nevertheless, under all above initial conditions, the NVI cost does converge to the true optimal cost $j_{\phi^\circ}^*$ obtained by the relative value iteration algorithm (with the true parameter values) presented in [18]. Note that after the transient period, the relative error of the NVI cost (with respect to $j_{\phi^\circ}^*$) reduces to less than 4×10^{-3} for $k \geq 50000$. The relative error in convergence of the NVI cost is illustrated in Fig. 7 for several initial estimates. The error values are computed from the NVI costs which are averaged over each 500 time steps. Note that after the early transition stage, the relative error reduces to less than 4×10^{-3} for $k \geq 50\,000$.

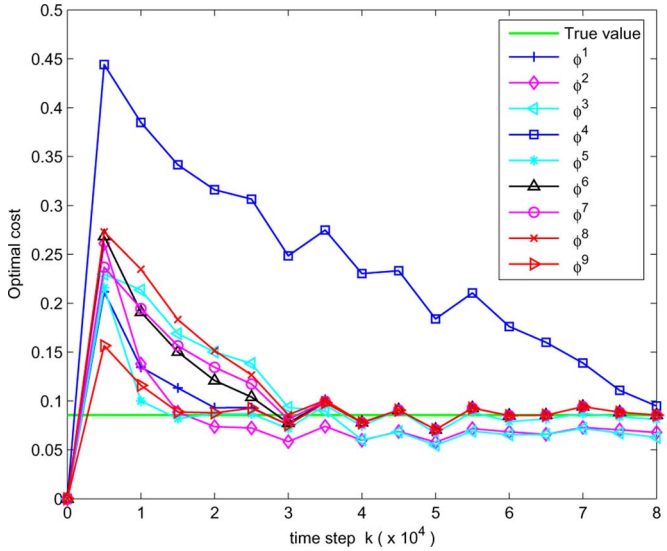


Fig. 6. Convergence of the NVI cost $j_{\hat{\phi}_k}^{\bar{\lambda}_k}$ under different initial conditions. The green line represents true optimal cost $j_{\phi^\circ}^*$.

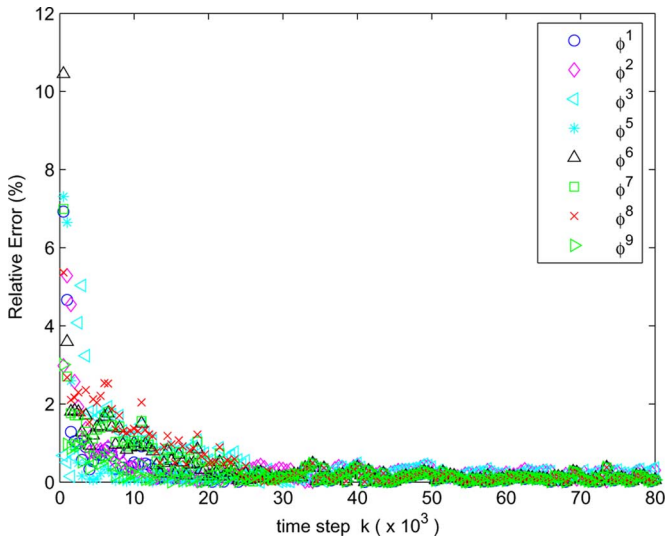


Fig. 7. Relative error between the NVI cost $j_{\hat{\phi}_k}^{\bar{\lambda}_k}$ and true optimal cost $j_{\phi^\circ}^*$.

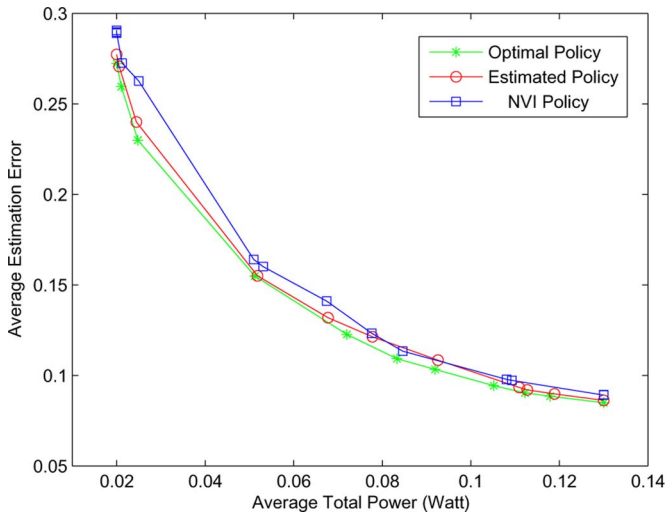


Fig. 8. Optimal and NVI Error/Power curves.

Fig. 8 shows the optimal state estimation error for various average sum power values across the sensors (obtained by varying the tradeoff factor β) for the NVI adaptive policy ($\bar{\lambda}_k$ as $k \to \infty$) and the optimal policy $\lambda_{\phi^\circ}^*$ (the relative value iteration algorithm of [18] for the true parameter values). As the available average sum power becomes low, the quality of the parameter estimates becomes poorer. As a result, the difference between the average state estimation error computed by the NVI policy and the one computed based on the optimal policy (with the knowledge of the true parameters) becomes larger. The solution to this is (if possible) to first compute the parameter estimates using the highest power levels at the sensors in order to learn the model parameters more accurately. This could be thought of as a training phase. Then, based on these estimated parameters, one may perform state estimation while jointly finding the optimal quantization and power allocation policies using the relative value iteration algorithm of [18] for a given average power constraint (see also [19]). The performance of this scheme is shown by the curve titled as the "Estimated Policy" in Fig. 8. As expected, the performance of this scheme is much closer to that of the optimal policy $\lambda_{\phi^\circ}^*$.

VII. Conclusion

In this paper, we have presented a novel method for jointly obtaining state and parameter estimates and designing optimal quantizer thresholds and transmit powers for a multi-sensor HMM estimation problem. The problem is motivated by monitoring applications of bandwidth and power constrained wireless sensor networks where multiple sensors observe an underlying Markov chain and binary quantize their noisy measurements before sending them to a remote fusion center over noisy wireless fading channels. The main contribution of this paper lies in proposing a novel, intelligent coupled recursive ML-based parameter estimation algorithm and a nonstationary value iteration (NVI)-based adaptive MDP control algorithm at the fusion center, for minimizing the expected state estimation error under an average sum power constraint across the sensors. Convergence of the parameter estimates and the asymptotic average optimality of the NVI algorithm are analytically proved (following existing results in [16] and [22]) under typical assumptions on the underlying Markov chain and the associated noise processes. Performance of this coupled algorithm in terms of the average state estimation error for various average sum power values are also illustrated via extensive numerical studies.

Finally, it must be said that although we focus on a particular dynamic resource allocation problem for multisensor state estimation of an HMM over fading channels in a wireless sensor network, the coupled NVI based adaptive control and recursive ML-based parameter estimation algorithm can be applied to many other problems where MDP or partially observed MDP-based control algorithms are used to make dynamic decisions but the parameters of the underlying processes are unknown. In fact, in most practical scenarios, the parameters of the underlying Markov chains, associated noise processes and fading channels (in the case of wireless networks) are not known *a priori* and need to be estimated in real-time. Examples of such MDP-based adaptive control applications abound in the literature on wireless networks for design of scheduling, transmission control, power control, admission control, and dynamic spectrum access algorithms as well as in other types of networks such as optical and satellite communication networks. References

are too many to be included here but an earlier survey of such applications to communication networks can be found in [50].

APPENDIX
SUMMARY OF THE ESTIMATION EQUATIONS

In the following, we summarize the estimation equations for our three different types of parameters based on the online algorithm presented in (12). Suppose $\mathbf{o}_{m,k+1} = (y_m^f, z_m, r_m, p_m)$ is the (incomplete) data observed at the fusion center from mth sensor at time $k+1$ and let $\mathbf{o}_{k+1} = (\mathbf{y}^f, \mathbf{z}, \mathbf{r}, \mathbf{p})$ be the observation vector received from all sensors. Also, let $\hat{\phi}_k = (\mathbf{S}(\hat{\phi}_k), \mathcal{S}(\hat{\phi}_k), \vartheta(\hat{\phi}_k))$ be the model estimate at time k. Then, $s_{ij}(\hat{\phi}_{k+1})$, and $\tilde{x}_i(\hat{\phi}_{k+1})$ for $i, j \in \{1, \ldots, n\}$ can be computed as follows:

$$s_{ij}(\hat{\phi}_{k+1}) = s_{ij}(\hat{\phi}_k) + 2\epsilon_{k+1}\left[\frac{\bar{\xi}_{k|k+1}(i,j)}{s_{ij}(\hat{\phi}_k)} - s_{ij}(\hat{\phi}_k)\sum_{j=1}^{n}\bar{\xi}_{k|k+1}(i,j)\right] \quad (23)$$

$$\tilde{x}_i(\hat{\phi}_{k+1}) = \tilde{x}_i(\hat{\phi}_k) + \epsilon_{k+1}\left[\frac{\bar{\gamma}_{k+1|k+1}(i)}{\psi_i(\mathbf{o}_{k+1}; \vartheta(\phi))} \times \frac{\partial}{\partial \tilde{x}_i(\phi)}\psi_i(\mathbf{o}_{k+1}; \vartheta(\phi))\right]_{\phi=\hat{\phi}_k} \quad (24)$$

where the state-to-observation probability mass function ψ_i is given by

$$\psi_i(\mathbf{o}_{k+1}; \vartheta(\phi)) = \sum_{\mathbf{y}^q}\left[\prod_{m=1}^{M} q_{i_m j_m}^m(z_m, p_m) F_i^m(y_m^q, r_m; \phi)\right]$$

in which $\mathbf{y}^q = (y_1^q, \ldots, y_M^q) \in \{b_1, b_2\}^M$, $y_m^q = b_{i_m}$, $y_m^f = b_{j_m}$ for $i_m, j_m \in \{1, 2\}$, and $F_i^m(y_m^q, r_m; \phi)$ is the state-to-observation probability mass function for sensor measurement y_m^q and is calculated as

$$F_i^m(y_m^q, r_m; \phi) = \mathbf{I}_{\{b_1\}}(y_m^q) + \frac{1}{2}\left(\mathbf{I}_{\{b_2\}}(y_m^q) - \mathbf{I}_{\{b_1\}}(y_m^q)\right) \times \operatorname{erfc}\left(\frac{r_m - \tilde{x}_i(\phi)}{\sqrt{2}(\vartheta^m(\phi))^2}\right) \quad (25)$$

where $\mathbf{I}_{\mathbf{A}}(\acute{e})$ is the indicator function on set $\mathbf{A}$ which takes the value 1 if $\acute{e} \in \mathbf{A}$ and 0 otherwise. The partial derivative of ψ_i with respect to the state level parameter $\tilde{x}_i(\phi)$ in (24) can be computed in a straightforward manner as follows:

$$\frac{\partial}{\partial \tilde{x}_i(\phi)}\psi_i(\mathbf{o}_{k+1}; \vartheta(\phi)) = \sum_{\mathbf{y}^q}\sum_{m=1}^{M}\left[q_{i_m j_m}^m(z_m, p_m) \times \left(\mathbf{I}_{\{b_2\}}(y_m^q) - \mathbf{I}_{\{b_1\}}(y_m^q)\right) \times \frac{1}{\sqrt{2\pi}(\vartheta^m(\phi))^2} e^{-(r_m - \tilde{x}_i(\phi))^2/2(\vartheta^m(\phi))^4} \times \prod_{\substack{l=1 \\ l\neq m}}^{M} q_{i_l j_l}^l(z_l, p_l) F_i^l(y_l^q, r_l; \phi)\right].$$

Finally, $\vartheta^m(\hat{\phi}_{k+1})$ is calculated as follows:

$$\vartheta^m(\hat{\phi}_{k+1}) = \vartheta^m(\hat{\phi}_k) + 2\epsilon_{k+1}\left[\sum_{i=1}^{n}\frac{\bar{\gamma}_{k+1|k+1}(i)}{\psi_i(\mathbf{o}_{k+1}; \vartheta(\phi))} \times \frac{\partial}{\partial \vartheta^m(\phi)}\psi_i(\mathbf{o}_{k+1}; \vartheta(\phi))\right]_{\phi=\hat{\phi}_k} \quad (26)$$

where the partial derivative of ψ_i with respect to $\vartheta^m(\phi)$ in (26) is given by

$$\frac{\partial}{\partial \vartheta^m(\phi)}\psi_i(\mathbf{o}_{k+1}; \vartheta(\phi)) = \sum_{\mathbf{y}^q}\sum_{m=1}^{M}\left[q_{i_m j_m}^m(z_m, p_m) \times \left(\mathbf{I}_{\{b_2\}}(y_m^q) - \mathbf{I}_{\{b_1\}}(y_m^q)\right) \times \frac{r_m - \tilde{x}_i(\phi)}{\sqrt{2\pi}(\vartheta^m(\phi))^3} e^{-(r_m - \tilde{x}_i(\phi))^2/2(\vartheta^m(\phi))^4} \times \prod_{\substack{l=1 \\ l\neq m}}^{M} q_{i_l j_l}^l(z_l, p_l) F_i^l(y_l^q, r_l; \phi)\right].$$

REFERENCES

[1] Z.-Q. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 735–744, Apr. 2005.

[2] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part II: Unknown probability density function," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2784–2796, Jul. 2006.

[3] E. Meschu, S. Roumeliotis, A. Ribeiro, and G. B. Giannakis, "Decentralized quantized Kalman filtering with scalable communication cost," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3727–3741, Aug. 2008.

[4] K. You, L. Xie, S. Sun, and W. Xiao, "Multiple-level quantized innovation kalman filter," in *Proc. IFAC World Congr. '08*, Seoul, Korea, Jul. 2008.

[5] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statist.*, vol. 37, no. 6, pp. 1554–1563, Dec. 1966.

[6] T. Petrie, "Probabilistic functions of finite state Markov chains," *Ann. Math. Statist.*, vol. 40, no. 1, pp. 97–115, Feb. 1969.

[7] B. G. Leroux, "Maximum-likelihood estimation for hidden Markov models," *Stoch. Processes their Applicat.*, vol. 40, pp. 127–143, 1992.

[8] P. J. Bickel and Y. Ritov, "Inference in hidden Markov models I: Local asymptotic normality in the stationary case," *Bernoulli*, vol. 2, pp. 199–228, 1996.

[9] U. Holst and G. Lindgren, "Recursive estimation in mixture models with Markov regime," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1683–1690, Nov. 1991.

[10] V. Krishnamurthy and J. B. Moore, "On-line estimation of hidden Markov model parameters based on the Kullback-Leibler information measure," *IEEE Trans. Signal Process.*, vol. 41, no. 8, pp. 2557–2573, Aug. 1993.

[11] T. Rydén, "On recursive estimation for hidden Markov models," *Stoch. Process. their Applicat.*, vol. 66, pp. 79–96, 1997.

[12] T. Rydén, "Asymptotically efficient recursive estimation for incomplete data models using the observed information," *Metrika*, vol. 44, pp. 119–145, 1998.

[13] V. Krishnamurthy, S. Dey, and J. P. LeBlanc, "Blind equalization of IIR channels using hidden Markov models and extended least squares," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2994–3006, Dec. 1995.

[14] S. Dey, V. Krishnamurthy, and T. Salmon-Legagneur, "Estimation of Markov-modulated time-series via EM algorithm," *IEEE Signal Process. Lett.*, vol. 1, no. 10, pp. 153–155, Oct. 1994.

[15] F. LeGland and L. Mevel, "Recursive estimation in hidden Markov models," in *Proc. 36th IEEE Conf. Decision Control*, San Diego, CA, Dec. 1997, pp. 3468–3473.

[16] V. Krishnamurthy and G. G. Yin, "Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 458–476, Feb. 2002.
[17] M. Huang and S. Dey, "Dynamic quantizer design for hidden Markov state estimation via multiple sensors with fusion center feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2887–2896, Aug. 2006.
[18] M. Huang and S. Dey, "Dynamic quantization for multisensor estimation over bandlimited fading channels," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4696–4702, Sep. 2007.
[19] N. Ghasemi and S. Dey, "Power-efficient dynamic quantization for multisensor HMM state estimation over fading channels," in *Proc. IEEE Int. Symp. Commun., Control, Signal Process. (ISCCSP'08)*, Mar. 12–14, 2008, pp. 1553–1558.
[20] N. Ghasemi and S. Dey, "A constrained MDP approach to dynamic quantizer design for HMM state estimation," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1203–1209, Mar. 2009.
[21] O. Hernández-Lerma, "Approximation and adaptive control of Markov processes: Average reward criterion," *J. Kybernetika (Prague)*, vol. 23, pp. 265–288, 1987.
[22] O. Hernández-Lerma, *Adaptive Markov Control Processes*, F. John, J. E. Marsden, and L. Sirovich, Eds. New York: Springer-Verlag, 1989.
[23] M. Kurano, "Discrete-time Markovian decision processes with an unknown parameter-Average return criterion," *J. Operat. Res. Soc. Japan*, vol. 15, pp. 67–76, 1972.
[24] P. Mandl, "Estimation and control in Markov chains," *J. Adv. Appl. Probab.*, vol. 6, pp. 40–60, Mar. 1974.
[25] V. Borkar and P. Varaiya, "Identification and adaptive control of Markov chains," *SIAM J. Control Optimiz.*, vol. 20, pp. 470–489, 1982.
[26] Y. M. El-Fattah, "Gradient approach for recursive estimation and control in finite Markov chains," *J. Adv. Appl. Probab.*, vol. 13, pp. 778–803, 1981.
[27] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, B. Rozovskii and M. Yor, Eds., 2nd ed. New York: Springer-Verlag, 2003, vol. 35, Applications of Mathematics.
[28] N. Ghasemi and S. Dey, "Adaptive quantizer design for HMM state and parameter estimation," in *Proc. 48th IEEE Conf. Decision Control (CDC'09)*, pp. 920–927.
[29] E. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, Sep. 1960.
[30] E. O. Elliot, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, Sep. 1963.
[31] M. Hassan, M. M. Krunz, and I. Matta, "Markov-based channel characterization for tractable performance analysis in wireless packet networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 821–831, May 2004.
[32] S. Cui, J.-J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.
[33] M. Kurano, J.-I. Nakagami, and Y. Huang, "Constrained Markov decision processes with compact state and action spaces: The average case," *Optimization*, vol. 48, no. 2, pp. 255–269, 2000.
[34] O. Hernandez-Lerma, J. Gonzalez-Hernandez, and R. R. Lopez-Martinez, "Constrained average cost Markov control processes in Borel spaces," *SIAM J. Control Optim.*, vol. 42, no. 2, pp. 442–468, 2003.
[35] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: Wiley, 1994.
[36] H. Yu and D. P. Bertsekas, "Discretized approximations for POMDP with average cost," in *Proc. 20th Conf. Uncertainty Artif. Intell.*, Banff, AB, Canada, 2004, pp. 619–627.
[37] D. P. Bertsekas, "Convergence of discretization procedures in dynamic programming," *IEEE Trans. Autom. Control*, vol. AC-20, no. 3, pp. 415–419, Jun. 1975.
[38] D. P. Bertsekas, *Dynamic Programming and Stochastic Control*. New York: Academic, 1976.
[39] A. Federgruen and P. J. Schweitzer, "Nonstationary Markov decision problems with converging parameters," *J. Optimiz. Theory Applicat.*, vol. 34, no. 2, pp. 207–241, Jun. 1981.
[40] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statist.*, vol. 41, no. 1, pp. 164–171, 1970.
[41] L. E. Baum and J. A. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bull. Amer. Math. Soc.*, vol. 73, no. 3, pp. 360–363, 1967.
[42] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc., Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
[43] G. J. McLachlan, *The EM Algorithm and Extensions*, V. Barnett, Ed. *et al.* New York: Wiley, 1997.
[44] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, A. V. Balakrishnan, I. Karatzas, and M. Yor, Eds. Berlin, Heidelberg, Germany: Springer-Verlag, 1990, vol. 22, Applications of Mathematics.
[45] F. LeGland and L. Mevel, "Exponential forgetting and geometric ergodicity in hidden Markov models," *Math. Control, Signals, Syst.*, vol. 13, pp. 63–93, 2000.
[46] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[47] L. Shue, B. D. O. Anderson, and S. Dey, "Exponential stability of filters and smoothers for hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 8, pp. 2180–2194, Aug. 1998.
[48] V. Krishnamurthy and T. Rydén, "Consistent estimation of linear and non-linear autoregressive models with Markov regime," *J. Time Series Anal.*, vol. 19, pp. 291–307, 1998.
[49] A. Goldsmith, *Wireless Communications*. New York: Cambridge Univ. Press, 2005.
[50] E. Altman, "Applications of Markov decision processes in communication networks," in *Handbook of Markov Decision Processes*, E. A. Feinberg and A. Shwartz, Eds. Norwell, MA: Kluwer, 2002, pp. 489–536.

Nader Ghasemi (S'08–M'11) was born in Tehran, Iran, in 1976. He received the B.E. and M.S. degrees in computer engineering and the M.S. degree in information technology from the University of Sydney, Sydney, Australia, in 1996, 1999, and 2006, respectively, and the Ph.D. degree in electrical engineering from the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia, in 2011.

He was an Academic Visiting Scholar in the Department of Electrical and Computer Engineering, University of Maryland, College Park, in 2010.

Dr. Ghasemi was the recipient of an Australian Special Postgraduate Fellowship from the Australian Research Council from 2007 to 2010. His current research interests include nonlinear estimation, stochastic control, signal processing for sensor networks, networked control systems, and distributed estimation.

Subhrakanti Dey (SM'06) was born in India, in 1968. He received the B.Tech. and M.Tech. degrees from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, in 1991 and 1993, respectively, and the Ph.D. degree from the Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, Australia, in 1996.

He has been with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia, since February 2000, where he is currently a Full Professor. From September 1995 to September 1997 and September 1998 to February 2000, he was a postdoctoral Research Fellow with the Department of Systems Engineering, Australian National University. From September 1997 to September 1998, he was a Post-Doctoral Research Associate with the Institute for Systems Research, University of Maryland, College Park. His current research interests include networked control systems, wireless communications and networks, signal processing for sensor networks, and stochastic and adaptive estimation and control.

Prof. Dey currently serves on the Editorial Board of *Elsevier Systems and Control Letters*. He was also an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.