A Monte Carlo scheme for node-specific inference over wireless sensor networks

Luca Martino*, Jorge Plata-Chaves†, Francisco Louzada*

* Institute of Mathematical Sciences and Computing, Universidade de São Paulo (ICMC-USP), Brazil.
† Department of Electrical Engineering-ESAT, STADIUS, KU Leuven, B-3001 Leuven, Belgium

***Abstract*—In this work, we design an efficient Monte Carlo scheme for a node-specific inference problem where a vector of global parameters and multiple vectors of local parameters are involved. This scenario often appears in inference problems over heterogeneous wireless sensor networks where each node performs observations dependent on a vector of global parameters as well as a vector of local parameters. The proposed scheme uses parallel local MCMC chains and then an importance sampling (IS) fusion step that leverages all the observations of all the nodes when estimating the global parameters. The resulting algorithm is simple and flexible. It can be easily applied iteratively, or extended in a sequential framework.**

Keywords: *Node-specific inference; Parallel MCMC; Importance sampling; Wireless sensor networks.*

I. Introduction

In the last decades, there has been a considerable research effort in the field of inference over networks, e.g.,[1]-[10]. Furthermore, Bayesian methods have become very popular in signal processing during the last years and, with them, Monte Carlo (MC) techniques that are often necessary for the implementation of optimal a posteriori estimators [11]-[13]. Indeed, MC methods are powerful tools for numerical inference and optimization [14]-[18]. They are very flexible techniques. The only requirement needs for applying an MC technique is to be able to evaluate point-wise the posterior probability density function (pdf) [11]-[12].

More recently, motivated by different applications such as cooperative spectrum sensing [19]-[20] or power system state estimation [21], special attention is being paid to more general inference problems in which the observation model of the different nodes depend on different but overlapping vectors of parameters. This kind of estimation problems, usually referred to as Node-Specific Parameter Estimation (NSPE) problems, has generally been addressed by distributed algorithms that rely on low-complex linear estimation techniques, e.g.,[19]-[25]. However, to the authors knowledge, very little is known about the application of MC techniques to design algorithms that can solve NSPE problems. To bridge this gap, we consider an inference problem where the measurements of each node depend on a vector of local parameters and a vector global parameters. To solve this problem, we propose a MC-based method that simultaneously estimate both the vector of global parameters as well as the different vectors of local parameters observed by the different noes. The proposed solution employs parallel MCMC algorithms for analyzing the local features of the network and an importance sampling (IS) fusion for obtaining complete estimators of the global parameters. Each MCMC method addresses a different target posterior function, obtained by considering a subset of observations. This approach also presents several computational benefits from a Monte Carlo point of view (as remarked exhaustively in Section III-A). For instance, the mixing of the MCMC methods is facilitated by the reduced number of measurements involved in the partial posterior, since this partial posterior distribution is implicitly *tempered* [26], [27]-[18]. Furthermore, several parallel or related schemes, proposed in literature, could be adapted for this framework [8]-[10] and [27]-[33]. Numerical simulations show the advantages of the proposed approach.

II. Problem statement

In this work, we are interested in making inference of the following variable of interest,

$$\mathbf{\Theta} = \begin{bmatrix} \mathbf{\Theta}^{(G)} \\ \mathbf{\Theta}^{(L)} \end{bmatrix} = \begin{bmatrix} \mathbf{x} \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_M \end{bmatrix} \in \mathcal{A} \subseteq \mathbb{R}^{D_\theta}, \tag{1}$$

composed by the vector $\mathbf{x} = [x_1, \ldots, x_{d_X}]^\top \in \mathbb{R}^{d_X}$ of global parameters and the vectors $\mathbf{v}_m = [v_{1,m}, \ldots, v_{d_m,m}]^\top \in \mathbb{R}^{d_m}$ of local parameters at node m, for $m = 1, \ldots, M$. The vector $\mathbf{\Theta}^{(G)} = \mathbf{x}$ gathers all the global parameters, whereas $\mathbf{\Theta}^{(L)} = [\mathbf{v}_1, \ldots, \mathbf{v}_M]^\top \in \mathbb{R}^L$ is formed by all the vectors of local parameters. We receive a set of d_Y measurements, $\mathcal{Y} = \{z_1, z_2, \ldots, z_{d_Y}\}$, with each $z_j \in \mathbb{R}$ (we assume z_j be scalar only for simplicity), related to the variable of interest $\mathbf{\Theta}$. We consider M disjoint subset of $\mathcal{Y}$, i.e., we can write $\mathcal{Y} = \mathbf{y}_{1:M} = \bigcup_{m=1}^{M} \mathbf{y}_m$ where $\mathbf{y}_m \cap \mathbf{y}_k = \emptyset$ with $m \neq k$ and with $\mathbf{y}_m$ denoting the observations performed by node m. We assume that the observations performed by two different nodes m and k depend on the same vector of global parameters $\mathbf{x}$. We also consider that the observations performed by each

This work has been supported by the Grant 2014/23160-6 of the São Paulo Research Foundation (FAPESP), by the Grant 305361/2013-3 of the National Council for Scientific and Technological Development (CNPq) and ERC Consolidator Grant SEDAL ERC-2014-CoG 647423. The project HANDiCAMS acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 323944.

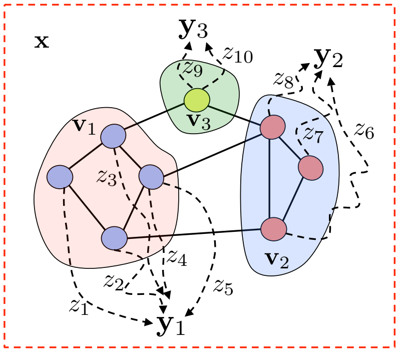

Figure 1. Wireless sensor network with node-specific observation models. In this graphical examples, the network is composed by 8 nodes, dividing in 3 clusters and providing 3 different vectors of measurements $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$. In this case, $\mathcal{Y} = \mathbf{y}_1 \cup \mathbf{y}_2 \cup \mathbf{y}_3$, $\mathbf{y}_j \cap \mathbf{y}_k \neq 0$ for $j \neq k$ (in this case, $d_{Y_1} = 5$, $d_{Y_2} = 3$, $d_{Y_3} = 2$). Each sensor can provide more than one measurement.

node m, i.e., $\mathbf{y}_m$, depends on the vector of local parameters $\mathbf{v}_m$. This setting appears naturally in several applications, for instance in the so-called NSPE problem over a wireless sensor network, e.g., ,[19]-[22] and [24]-[25] (see Figure 1). Furthermore, in the considered setting we assume that the observations performed by different nodes are conditionally independent, i.e., the likelihood function can be factorized as

$$L(\mathbf{y}_{1:M}|\boldsymbol{\Theta}) \quad = \quad \prod_{m=1}^{M} \ell_m(\mathbf{y}_m|\mathbf{x}, \mathbf{v}_m). \tag{2}$$

Considering a prior probability density function (pdf) $p(\boldsymbol{\Theta}) = p(\mathbf{x}) \prod_{m=1}^{M} p(\mathbf{v}_m)$, the *complete posterior* pdf can written as

$$\Omega(\boldsymbol{\Theta}|\mathbf{y}_{1:M}) = \prod_{m=1}^{M} \bar{\pi}_m(\mathbf{x}, \mathbf{v}_m|\mathbf{y}_m), \tag{3}$$

where the *partial posteriors* are

$$\begin{aligned} &\bar{\pi}_m(\mathbf{x}, \mathbf{v}_m|\mathbf{y}_m) \propto \pi_m(\mathbf{x}, \mathbf{v}_m|\mathbf{y}_m) \quad \text{where} \\ &\pi_m(\mathbf{x}, \mathbf{v}_m|\mathbf{y}_m) = \ell_m(\mathbf{y}_m|\mathbf{x}, \mathbf{v}_m) \left[p(\mathbf{x})\right]^{1/M} p(\mathbf{v}_m). \end{aligned} \tag{4}$$

For the inference of $\mathbf{x}$, an important role is played by the marginal posterior density

$$\begin{aligned} G(\mathbf{x}|\mathbf{y}_{1:M}) \quad &= \int_{\mathbb{R}^L} \Omega(\boldsymbol{\Theta}|\mathbf{y}_{1:M}) d\mathbf{v}_1 \ldots d\mathbf{v}_M & (5) \\ &= \int_{\mathbb{R}^L} \prod_{m=1}^{M} \bar{\pi}_m(\mathbf{x}, \mathbf{v}_m|\mathbf{y}_m) d\mathbf{v}_1 \ldots d\mathbf{v}_M & (6) \\ &= \prod_{m=1}^{M} g_m(\mathbf{x}|\mathbf{y}_m), & (7) \end{aligned}$$

where

$$g_m(\mathbf{x}|\mathbf{y}_m) = \int_{\mathbb{R}^{d_m}} \bar{\pi}_m(\mathbf{x}, \mathbf{v}_m|\mathbf{y}_m) d\mathbf{v}_m. \tag{8}$$

Note that the integrals above cannot be computed analytically, in general. Furthermore, notice that a similar approach is considered in the Big Data context, e.g., [8]-[10] and [33].

III. Bayesian Inference

Our purpose is to infer $\boldsymbol{\Theta} = [\mathbf{x}, \mathbf{v}_{1:M}]$ given $\mathcal{Y} = \mathbf{y}_{1:M}$. For instance, we desire to compute the *Minimum Mean Square Error* (MMSE) estimators of $\mathbf{x}$ and $\mathbf{v}_{1:M}$, i.e.,

$$\begin{aligned} \widehat{\mathbf{x}} \quad &= \int_{\mathcal{A}} \mathbf{x} \ \Omega(\mathbf{x}, \mathbf{v}_{1:M}|\mathbf{y}_{1:M}) d\mathbf{x} d\mathbf{v}_{1:M}, & (9) \\ &= \int_{\mathbb{R}^{d_X}} \mathbf{x} \ G(\mathbf{x}|\mathbf{y}_{1:M}) d\mathbf{x}, & (10) \end{aligned}$$

and

$$\widehat{\mathbf{v}}_m \quad = \int_{\mathcal{A}} \mathbf{v}_m \ \Omega(\mathbf{x}, \mathbf{v}_{1:M}|\mathbf{y}_{1:M}) d\mathbf{x} d\mathbf{v}_{1:M}, \tag{11}$$

for $m = 1, \ldots, M$. In general, we are not able to calculate analytically the integrals above. Thus, we apply a Monte Carlo (MC) approach for computing approximately $\widehat{\mathbf{x}}$ and $\widehat{\mathbf{v}}_{1:M}$.

A. Benefits of the parallel MC implementation

The previous factorization of the posterior pdf suggests the use of M parallel algorithms and then combine the corresponding outputs. This is convenient from a Monte Carlo point of view. Namely, the use of M parallel MC methods, each one addressing one partial posterior $\bar{\pi}_m$, presents several computational benefits:

a) Each partial posterior $\bar{\pi}_m$ is embedded in a state space of lower dimensionality, specifically, $\mathbf{x}, \mathbf{v}_m \in \mathbb{R}^{d_X \times d_m}$. This clearly helps the exploration of the space by the MC algorithm.

b) Each partial posterior involves a smaller number of measurements. This is an advantage since the mass of probability is in general more disperse than when a big number of observations is jointly considered, producing a *tempering effect* (data-tempering) [18], [26], [28]. This again helps the exploration of the state space (as suggested, e.g., in [28]).

c) This scenario automatically allows a possible parallel implementation over a wireless sensor network where each node in the network undertakes one of the MC methods.

IV. Monte Carlo Node-Specific Inference

Let us consider that we able to draw N samples $\boldsymbol{\theta}_m^{(n)} = [\mathbf{x}^{(n)}, \mathbf{v}_m^{(n)}]$, with $n = 1, \ldots, N$, directly from each $\bar{\pi}_m$, i.e.,

$$\boldsymbol{\theta}_m^{(1)}, \ldots, \boldsymbol{\theta}_m^{(N)} \sim \bar{\pi}_m(\mathbf{x}, \mathbf{v}_m|\mathbf{y}_m),$$

with $m = 1, \ldots, M$. Due to the factorization of the complete target pdf $\Omega(\boldsymbol{\Theta}|\mathbf{y}_{1:M})$, for inferring the local variable $\mathbf{v}_m$, we use only the samples $\mathbf{v}_m^{(n)}$, $n = 1, \ldots, N$, obtained from $\bar{\pi}_m$. For the global variable $\mathbf{x}$, we can build M different partial Monte Carlo estimators. However, all the information contained in the different partial posteriors should be employed for providing a more efficient unique estimator of $\mathbf{x}$. This can be done by combining adequately the M partial Monte Carlo estimators of $\mathbf{x}$.

Let us consider that we are able to draw samples from each marginal pdfs $g_m(\mathbf{x}|\mathbf{y}_m)$ in Eq. (8) and also assume that we are able to evaluate it. In this case, we can use the following IS scheme:

1) Draw $\mathbf{x}_m^{(1)}, \ldots, \mathbf{x}_m^{(N)} \sim g_m(\mathbf{x}|\mathbf{y}_m)$.

2) Assign the weight

$$w_m^{(n)} = \frac{G(\mathbf{x}_m^{(n)}|\mathbf{y}_{1:M})}{g_m(\mathbf{x}_m^{(n)}|\mathbf{y}_m)} = \prod_{k=1;k\neq m}^{M} g_k(\mathbf{x}_k^{(n)}|\mathbf{y}_k), \quad (12)$$

to each sample $\mathbf{x}_m^{(n)}$, for $n = 1, \dots, N$.

Then, the IS approximation of the MMSE estimator $\widehat{\mathbf{x}}$ is

$$\widetilde{\mathbf{x}} = \frac{1}{\sum_{m=1}^{M}\sum_{n=1}^{N} w_m^{(n)}} \sum_{m=1}^{M}\sum_{n=1}^{N} w_m^{(n)}\mathbf{x}_m^{(n)} \quad (13)$$

However, the previous approach has two main problems:

- We are not able to draw from $g_m(\mathbf{x}|\mathbf{y}_m)$ in Eq. (8).
- It is not possible to evaluate $g_m(\mathbf{x}|\mathbf{y}_m)$ and $G(\mathbf{x}_m|\mathbf{y}_{1:M})$.

Next, we will show how the previous problems can be overcome.

A. Proposed Algorithm

A possible approximate solution consists in the following procedure. First, we use parallel MCMC algorithms to draw samples $\boldsymbol{\theta}_m^{(n)} = [\mathbf{x}^{(n)}, \mathbf{v}_m^{(n)}]$ from each partial posterior $\bar{\pi}_m(\mathbf{x}, \mathbf{v}_m|\mathbf{y}_m)$. Given an index $m \in \{1, \dots, M\}$, once the chain has converged, note that the samples $\mathbf{x}_m^{(n)}$ are distributed as $g_m(\mathbf{x}|\mathbf{y}_m)$ whereas $\boldsymbol{\theta}_m^{(n)}$ is distributed as $\bar{\pi}_m(\mathbf{x}, \mathbf{v}_m|\mathbf{y}_m)$. Then, we build a Kernel Density Estimator (KDE) [34],

$$\widehat{g}_m(\mathbf{x}|\mathbf{y}_m) = \frac{1}{N}\sum_{n=1}^{N} \varphi(\mathbf{x}|\mathbf{x}_m^{(n)}, \mathbf{C}), \quad (14)$$

for each $m \in \{1, \dots, M\}$, using the generated samples $\mathbf{x}_m^{(n)}$, $n = 1, \dots, N$, as means of a kernel function φ with bandwidth matrix $\mathbf{C}$. In this way, we can compute an approximate weight

$$\widehat{w}_m^{(n)} = \prod_{k=1;k\neq m}^{M} \widehat{g}_k(\mathbf{x}_k^{(n)}|\mathbf{y}_k), \quad (15)$$

so that the IS estimator $\widetilde{\mathbf{x}}$ in Eq. (13) can be approximated with

$$\bar{\mathbf{x}} = \frac{1}{\sum_{m=1}^{M}\sum_{n=1}^{N} \widehat{w}_m^{(n)}} \sum_{m=1}^{M}\sum_{n=1}^{N} \widehat{w}_m^{(n)}\mathbf{x}_m^{(n)}. \quad (16)$$

In this way, we obtain the Parallel Marginal Markov Importance Sampler (PMMIS) for node-specific inference summarized in Table I and Figure 2. Moreover, Table II summarizes the notation about the estimators of $\mathbf{x}$.

Theoretical support. After a burn-in period, the MCMC chains converges to the invariant target pdf (for $N \to \infty$, the convergence is ensured [11], [12]). Namely, after some iterations, the MCMC methods yields samples $\{\mathbf{x}_m^{(n)}, \mathbf{v}_m^{(n)}\}$ distributed according to $\bar{\pi}_m$, so that $\{\mathbf{x}_m^{(n)}\}$ are distributed as the marginal partial posteriors $g_m(\mathbf{x}|\mathbf{y}_m)$ [11]. There exists an optimal bandwidth $\mathbf{C}^*$ [34] such that

$$\widehat{g}_m(\mathbf{x}|\mathbf{y}_m) \to g_m(\mathbf{x}|\mathbf{y}_m), \qquad \text{for} \qquad N \to \infty.$$

As a consequence, we have that $\widehat{w}_m^{(n)} \to w_m^{(n)}$ and $\bar{\mathbf{x}} \to \widetilde{\mathbf{x}}$, for $N \to \infty$. Moreover, the IS estimator is consistent [11] so that $\bar{\mathbf{x}} \to \widetilde{\mathbf{x}} \to \widehat{\mathbf{x}}$, as $N \to \infty$. In general, the optimal bandwidth $\mathbf{C}^*$ is unknown. However, using a bandwidth $\mathbf{C} \neq \mathbf{C}^*$, Eq. (14) provides an estimator of $g_m(\mathbf{x}|\mathbf{y}_m)$ [34], in any case.

Table I
PARALLEL MARGINAL MARKOV IMPORTANCE SAMPLER (PMMIS).

1. **Local MCMC samplers:** Generate M chains of length N, i.e.,
$$\boldsymbol{\theta}_m^{(1)} = [\mathbf{x}^{(1)}, \mathbf{v}_m^{(1)}], \dots, \boldsymbol{\theta}_m^{(N)} = [\mathbf{x}^{(N)}, \mathbf{v}_m^{(N)}],$$
with target pdf $\bar{\pi}_m(\mathbf{x}|\mathbf{y}_m) \propto \pi_m(\mathbf{x}|\mathbf{y}_m)$, $m = 1, \dots, M$.
2. **Kernel density estimation:** Build
$$\widehat{g}_m(\mathbf{x}|\mathbf{y}_m) = \frac{1}{N}\sum_{n=1}^{N} \varphi(\mathbf{x}|\mathbf{x}_m^{(n)}, \mathbf{C}), \quad (17)$$
given a function φ and scale parameter $\mathbf{C}$.
3. **Global IS fusion:** Compute the weights
$$\widehat{w}_m^{(n)} = \prod_{k=1;k\neq m}^{M} \widehat{g}_k(\mathbf{x}_k^{(n)}|\mathbf{y}_k), \quad (18)$$
for $m = 1, \dots, M$ and $n = 1, \dots, N$.
4. **Outputs:** Return the Monte Carlo estimators
$$\bar{\mathbf{x}} = \frac{1}{\sum_{m=1}^{M}\sum_{n=1}^{N} \widehat{w}_m^{(n)}} \sum_{m=1}^{M}\sum_{n=1}^{N} \widehat{w}_m^{(n)}\mathbf{x}_m^{(n)}, \quad (19)$$
$$\widetilde{\mathbf{v}}_m = \frac{1}{N}\sum_{n=1}^{N} \mathbf{v}_m^{(n)}, \quad m = 1, \dots, M. \quad (20)$$

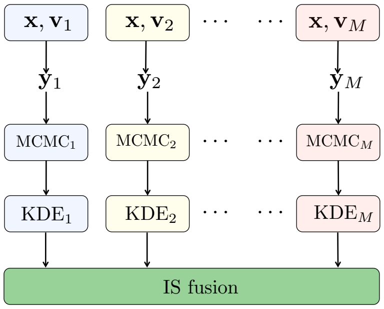

Figure 2. Graphical representation of PMMIS scheme: M local MCMC chains are run, each one addressing one partial posterior $\bar{\pi}_m$. A KDE for approximating the marginal partial posteriors $g_m(\mathbf{x}|\mathbf{y}_m)$ is performed locally. Then, an IS fusion is employed to estimate the global parameter $\mathbf{x}$.

Table II
NOTATION OF DIFFERENT ESTIMATORS OF $\mathbf{x}$.

Notation	Description
$\widehat{\mathbf{x}}$	MMSE estimator in Eq. (9).
$\widetilde{\mathbf{x}}$	Monte Carlo estimator (approximation of $\widehat{\mathbf{x}}$) in Eq. (13).
$\bar{\mathbf{x}}$	Approximate Monte Carlo estimator (approximation of $\widetilde{\mathbf{x}}$) in Eqs. (16)-(19).

Alternative IS weights. Other *proper* IS weights can be employed in our framework, providing consistent estimators [35]-[37]. For instance, a full deterministic mixture approach [36], [38]-[39] for multiple importance sampling (MIS) schemes can be used, i.e.,

$$w_m^{(n)} = \frac{G(\mathbf{x}_m^{(n)}|\mathbf{y}_{1:M})}{\frac{1}{M}\sum_{j=1}^{M} g_j(\mathbf{x}_m^{(n)}|\mathbf{y}_j)} = \frac{\prod_{k=1}^{M} g_k(\mathbf{x}_m^{(n)}|\mathbf{y}_k)}{\frac{1}{M}\sum_{k=1}^{M} g_k(\mathbf{x}_m^{(n)}|\mathbf{y}_k)}. \quad (21)$$

It is possible to show that the application of these DM-MIS weights provides more efficient IS estimators [36], [37] (i.e., with smaller variance).

B. Parallel Metropolis-Hastings algorithms

For simplicity, we consider Metropolis-Hastings (MH) methods in the first step of the novel scheme. However, more sophisticated algorithms can be employed. More specifically, starting with a randomly chosen $\boldsymbol{\theta}_m^{(0)}$, we perform the following steps:

```
For $m = 1, \ldots, M$:
  For $n = 1, \ldots, N$:
    1) Draw $\boldsymbol{\theta}'$ from a proposal pdf $q_m(\boldsymbol{\theta}'|\boldsymbol{\theta}_m^{(n-1)})$.
    2) Set $\boldsymbol{\theta}_m^{(n)} = \boldsymbol{\theta}'$, with probability
           $\alpha = \min \left[ 1, \frac{\pi_m(\boldsymbol{\theta}'|\mathbf{y}_m) q_m(\boldsymbol{\theta}_m^{(n-1)}|\boldsymbol{\theta}')}{\pi_m(\boldsymbol{\theta}_m^{(n-1)}|\mathbf{y}_m) q_m(\boldsymbol{\theta}'|\boldsymbol{\theta}_m^{(n-1)})} \right], \quad (22)$
       otherwise set $\boldsymbol{\theta}_m^{(n)} = \boldsymbol{\theta}_m^{(n-1)}$ (with probability $1 - \alpha$).
```

V. Numerical simulations

In order to test PMMIS, we consider a Gaussian likelihoods

$$f_m(\mathbf{y}_m|x, v_m) = \prod_{j=1}^{d_{Y_m}} \mathcal{N}(z_j|x, v_m, \boldsymbol{\Sigma}_m), \tag{23}$$

with $\mathbf{y}_m = [z_1, \ldots, z_{d_{Y_m}}]^\top$, $m = 1, \ldots, M$, and where $z_j, x, v_m \in \mathbb{R}$. Note that $d_{Y_m} = |\mathbf{y}_m|$. We consider flat improper priors over x and v_m, $m = 1, \ldots, M$. We set $M = 10$, so that we have 10 different partial target pdfs $\bar{\pi}_m(x, v_m|\mathbf{y}_m) \propto f_m(\mathbf{y}_m|x, v_m)$. The covariance matrices are $\boldsymbol{\Sigma}_m = [\sigma_{1,m}^2 \; \rho_m; \rho_m \; \sigma_{2,m}^2]$ with ,

$$\begin{aligned}
\sigma_{1,1:M} &= \left[\frac{1}{2}, \frac{3}{2}, 4, \frac{5}{2}, 3, \frac{7}{2}, 3, \frac{5}{2}, 2, \frac{1}{2}\right], \\
\sigma_{2,1:M} &= \left[\frac{1}{3}, \frac{2}{3}, 1, \frac{4}{3}, \frac{5}{3}, 2, \frac{7}{3}, \frac{8}{3}, 3, \frac{10}{3}\right], \\
\rho_{1:M} &= \left[0, \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}, \frac{5}{10}, \frac{6}{10}, \frac{7}{10}, \frac{8}{10}, \frac{9}{10}\right]
\end{aligned}$$

We set $x = -1$ and $v_{1:M} = [-5, -4, -3, -2, -1, 0, 1, 2, 3, 4]$ as "true" values of the parameters. Then, given these values, we generate (according to the model in Eq. (23)) different numbers of measurements for each partial likelihood, specifically, $d_{Y_{1:M}} = [2, 2, 50, 2, 5, 20, 5, 100, 2, 10]$. Given a set $\mathcal{Y} = \mathbf{y}_{1:M}$ of generated observations, in this toy example we can compute the MMSE estimator $\widehat{x} = -0.911$ by a costly deterministic numerical procedure using a thin grid (for approximating the marginal posterior and then the corresponding expected value). Thus, we apply PMMIS in 400 independent runs and compare the obtained estimator $\bar{x}$ with $\widehat{x} = -0.911$, computing the corresponding MSE.

We consider Gaussian functions φ for the KDE, with the optimal bandwidth suggested in [34]. For the proposal pdfs q_m's of the MH algorithms, we employ standard Gaussian random walk proposals (with identity covariance matrix [1 0; 0 1]). We test PMMIS for different values of the length of the chains, N, from $N = 15$ to $N = 2000$. Furthermore, at each run, we also compute a trivial Monte Carlo approximation of $\widehat{x}$, given by

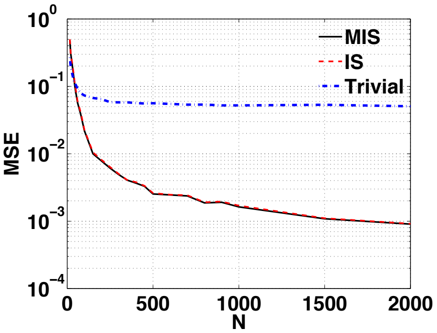

Figure 3. MSE in the estimation of $\hat{x}$ obtaining by the Monte Carlo approximation $\bar{x}$ by PMMIS using standard IS weights (dashed line), DM-MIS weights (solid line) and the trivial weights $\frac{1}{MN}$ (dotted-dashed line), for the final fusion (semilog scale representation).

$$\bar{x}_{trivial} = \frac{1}{M} \sum_{m=1}^{M} \tilde{x}_m = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} x_m^{(n)},$$

where $\tilde{x}_m = \frac{1}{N} \sum_{n=1}^{N} x_m^{(n)}$ is the Monte Carlo approximation of $\widehat{x}_m$, obtained using the samples of m-th chain. The results are shown in Figure 3, in terms of MSE versus N (length of the chains, i.e., number of MH iterations for each partial target). Three curves are shown, corresponding to the use of standard IS weights (dashed line), multiple IS weights (solid line) and the trivial solution (dotted-dashed line). We can observe that PMMIS provides good results outperforming the trivial solution for $N > 40$. This means that, for $N \leq 40$, the samples generated by the MH methods still belong to the "burn-in" period and the convergence to the invariant pdf is not reached. However, with a adequate number of iterations of the chains, PMMIS provides good results. In the example, the standard IS and DM-MIS weights perform similarly (with a slight advantages for the DM-MIS weights).

VI. Conclusions

In this work, we have introduced a novel Monte Carlo scheme in order to obtain an efficient solution for a node-specific inference problem. The new Bayesian method provides a simultaneous estimation of local and global parameters. The estimation of the global parameters takes into account all the possible statistical information. The proposed algorithm is an importance sampler that assigns weights to the samples obtained by the application of parallel MCMC methods. Each MCMC addresses a different partial target distribution, considering only a subset of measurements. As future line, we consider the possible design of an iterative implementation of the proposed scheme, where the proposal pdfs employed by the MCMC algorithms are adapted online, generating in this way an interaction among the parallel chains.

REFERENCES

[1] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.

[2] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, 2007.

[3] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1865–1877, 2008.

[4] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4583–4588, 2009.

[5] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2010.

[6] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks - part I: Sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, 2010.

[7] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Adaptive robust distributed learning in diffusion sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4692–4707, 2011.

[8] Steven L. Scott, Alexander W. Blocker, Fernando V. Bonassi, Hugh A. Chipman, Edward I. George, and Robert E. McCulloch, "Bayes and big data: The consensus Monte Carlo algorithm," in *EFaBBayes 250th conference*, 2013, vol. 16.

[9] W. Neiswanger, C. Wang, and E. Xing, "Asymptotically exact, embarrassingly parallel MCMC," *arXiv:1311.4780*, pp. 1–16, 21 Mar. 2014.

[10] R. Bardenet, A. Doucet, and C. Holmes, "On Markov chain Monte Carlo methods for tall data," *arXiv:1505.02827*, 2015.

[11] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.

[12] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2004.

[13] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, New York (USA), 2001.

[14] L. Martino and J. Míguez, "A generalization of the adaptive rejection sampling algorithm," *Statistics and Computing*, vol. 21, no. 11, pp. 633–647, July 2011.

[15] W. J. Fitzgerald, "Markov chain Monte Carlo methods with applications to signal processing," *Signal Processing*, vol. 81, no. 1, pp. 3–18, January 2001.

[16] M. Hong, M F Bugallo, and P M Djuric, "Joint model selection and parameter estimation by population monte carlo simulation," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 3, pp. 526–539, 2010.

[17] P. M. Djurić, B. Shen, and M. F. Bugallo, "Population Monte Carlo methodology a la Gibbs sampling," in *EUSIPCO*, 2011.

[18] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.

[19] N. Bogdanovic, J. Plata-Chaves, and K. Berberidis, "Distributed incremental-based LMS for node-specific adaptive parameter estimation," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5382–5397, 2014.

[20] J. Plata-Chaves, N. Bogdanovic, and K. Berberidis, "Distributed diffusion-based LMS for node-specific parameter estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 13, no. 63, pp. 3448–3460, 2015.

[21] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1617–1626, 2013.

[22] J. Plata-Chaves, N. Bogdanovic, and K. Berberidis, "Distributed incremental-based RLS for node-specific parameter estimation over adaptive networks," in *IEEE 21st European Signal Conference, 2013. EUSIPCO 2013*, 2013.

[23] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4129–4144, 2014.

[24] J. Chen, S. K. Ting, C. Richard, and A. H. Sayed, "Group diffusion LMS," in *IEEE 41th International Conference on Acoustics, Speech and Signal Processing, 2016. ICASSP 2016*, 2016.

[25] J. Plata-Chaves, M. H. Bahari, M. Moonen, and A. Bertrand, "Unsupervised diffusion-based LMS for node-specific parameter estimation over wireless sensor networks," in *IEEE 41th International Conference on Acoustics, Speech and Signal Processing, 2016. ICASSP 2016*, 2016.

[26] E. Marinari and G. Parisi, "Simulated tempering: a new Monte Carlo scheme," *Europhysics Letters*, vol. 19, no. 6, pp. 451–458, July 1992.

[27] A. Jasra, D. A. Stephens, and C. C. Holmes, "On population-based simulation for static inference," *Statistics and Computing*, vol. 17, no. 3, pp. 263–279, 2007.

[28] N. Chopin, "A sequential particle filter method for static models," *Biometrika*, vol. 89, no. 3, pp. 539–551, 2002.

[29] L. Martino, V. Elvira, D. Luengo, and J. Corander, "Layered adaptive importance sampling," *Statistics and Computing, to appear*, 2016.

[30] L. Martino, V. Elvira, D. Luengo, A. Artes, and J. Corander, "Orthogonal MCMC algorithms," *IEEE Statistical Signal Processing Workshop, (SSP)*, 2014.

[31] L. Martino, V. Elvira, D. Luengo, A. Artes, and J. Corander, "Smelly parallel MCMC chains," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015.

[32] A. Jasra, D. A. Stephens, and C. C. Holmes, "Population-based reversible jump Markov chain Monte Carlo," *Biometrika*, vol. 94, no. 4, pp. 787–807, 2007.

[33] D. Luengo, L. Martino, V. Elvira, and M. Bugallo, "Efficient linear combination of partial Monte Carlo estimators," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015.

[34] M.P. Wand and M.C. Jones, *Kernel Smoothing*, Chapman and Hall, 1994.

[35] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Efficient multiple importance sampling estimators," *Signal Processing Letters, IEEE*, vol. 22, no. 10, pp. 1757–1761, 2015.

[36] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized multiple importance sampling," *arXiv preprint arXiv:1511.03095*, 2015.

[37] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler: Learning from the uncertanity," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4422–4437, 2015.

[38] E. Veach and L. Guibas, "Optimally combining sampling techniques for Monte Carlo rendering," *In SIGGRAPH 1995 Proceedings*, pp. 419–428, 1995.

[39] A. Owen and Y. Zhou, "Safe and effective importance sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 135–143, 2000.