Incremental Distributed Identification of Markov Random Field Models in Wireless Sensor Networks

Anand Oka and Lutz Lampe
Dept. of Electrical and Computer Engg., Univ. of British Columbia, Canada.

Abstract

Wireless Sensor Networks (WSNs) comprise of highly power constrained nodes that observe a hidden natural field and reconstruct it at a distant data fusion center. Algorithmic strategies for extending the lifetime of such networks invariably require a knowledge of the statistical model of the underlying field. Since centralized model identification is communication intensive and eats into any potential power savings, we present a stochastic recursive identification algorithm which can be implemented in a fully distributed and scalable manner within the network. We demonstrate that it consumes modest resources relative to centralized estimation, and is stable, unbiased, and asymptotically efficient.

I. INTRODUCTION

A Wireless Sensor Network (WSN) is an ad hoc network of battery powered motes, which measures a physical field like the presence of a toxic chemical [1], and reconstructs it at a distant Fusion Center (FC). The transmission of all raw data out of the network is an energy-intensive procedure that quickly drains the batteries of the motes, and severely curtails the lifetime of the WSN [1]. Hence much research has been recently directed towards the exploitation of the spatiotemporal dependencies in the sensor data to improve the power efficiency of the network via strategies like Distributed Source Coding [2], Correlated Data Gathering [3], Source-Channel Decoding [4], Distributed Detection [5], [6], Distributed Filtering [7], Distributed Learning [8], and Energy Aware Routing [9]. In every case, some knowledge of the statistical model of the field is a prerequisite, e.g. to design optimal codes [2] or routing tables [9].

Since Maximum Likelihood (ML) identification via the Expectation Maximization (EM) algorithm of [10] is typically intractable for large models, low complexity stochastic recursive algorithms have been proposed in literature, which also have the advantage that they can seamlessly track model variations. For...
example, [11] has proposed an incremental EM algorithm for finite mixture models, [12] and [13] have proposed stochastic recursive estimators for the parameters of Hidden Markov Chain Models (HMMs), and [14]–[16] have proposed stochastic EM algorithms (SAEM) for spatial Hidden Markov Random Fields (HMRFs) in the context of image processing. Moreover, [13] and [14]–[16] use a Markov Chain ‘Monte-Carlo’ (MCMC) approximation [17] for their stochastic innovations, since an exact calculation is infeasible for large cyclical MRFs. A nice overview of incremental EM algorithms can be found in [18].

However, all the above noted researchers [11]–[16], [18] have exclusively considered centralized identification schemes, since there is no incentive for distributed algorithms in applications like image processing. In the WSN scenario on the other hand, transmission of all raw observation data to the FC for centralized identification requires energy-intensive communication, which greatly diminishes the net energy savings possible from the knowledge of the model. Therefore, it is clearly advantageous if the model identification can be done in situ, i.e. within the network itself. Any such proposal, however, must also satisfy the special constraints of the motes, and should be scalable with the size of the network, which makes a distributed algorithm mandatory. This novel problem of distributed in-situ model identification has received only a limited attention to date, e.g. identification of linear models [19], [20], incremental EM identification of fully observed mixture models [21], and supervised identification of MRF models [5]. To our knowledge, the practical problem of blind identification of hidden nonlinear probabilistic field models like HMRFs, using a spatially distributed algorithm, seems to be largely unexplored.

Contributions and organization of this paper: In this manuscript we address the above mentioned gap in literature by proposing a distributed incremental identification algorithm for a class of probabilistic models often encountered in WSN applications, namely, HMRFs [5], [22]. In Section II we first review the definition of such models and motivate their use, and point out which specific properties of HMRFs are prerequisites for the practicality of our identification procedure. We then specify our proposed incremental identification algorithm in Section III. While it is similar to the SAEM algorithm of [14]–[16] in that it makes a stochastic gradient ascent of the observation log-likelihood using MCMC techniques, we analyze our algorithm w.r.t. four issues crucial in a distributed implementation, namely (i) stability (Section III-A), (ii) covariance efficiency (Section III-B), (iii) scalability and power efficiency (Section IV), and (iv) robustness to early termination of MCMC iterations (Section V). Section VI concludes the paper.

Notation: \( \mathcal{N}(y|\mu, \Sigma) \) will denote a multivariate Gaussian p.d.f. with mean \( \mu \) and covariance \( \Sigma \). \( \mathbb{E}_{P(z|\gamma)}[f(Z)] \) will indicate an expectation of \( f(Z) \), w.r.t. probability mass/density function \( P(z|\gamma) \), with \( \gamma \) a parameter. If \( x \) is an \( N \)-vector, \( \sum_{x,z} \) will indicate a summation over all configurations of \( x = (x_1, x_2, \ldots, x_N) \), while \( \sum_{x, z_s} \) will indicate a summation over all configurations having the \( s^{th} \) component pinned to \( x_s \). In general,
when $A$ is a vector, $A_i$ is its $i^{th}$ component. $A^T$ is the transpose of a vector/matrix $A$. $\|A\|$ will signify the spectral norm of a square matrix $A$, and the Euclidean norm of a vector $A$. $\|A\|_F$ is the Frobenius norm of matrix $A$. $I_{N \times N}$ is the identity matrix of size $N$, and $0_N$ is an all-zero column vector of size $N$.

II. SYSTEM MODEL

A. Motivation for Identifying MRF Models

MRFs/HMRFs constitute a major class of parametric models based on graphs, and are popular for modelling a wide variety of random fields that are observed by WSNs. Specifically, Gauss-Markov random fields (GMRFs) are often used for modelling continuous valued spatio-temporal processes [3], [9], [22]–[24], with covariance structures chosen from classes such as spherical, power exponential, rational quadratic, Matérn etc, as befits the application [9]. On the other hand, a WSN may also be used to simply detect discrete valued variables like the presence or absence of an effluent, and subsequently to delineate the boundary of such a discrete phenomena. In such cases, a binary valued MRF, also known as a Boltzmann or Ising model, is typically used [5], [6], [25]. A third application involves the use of Boltzmann/Ising models (already popular in centralized image processing [14]–[16], [26]) for the distributed detection of textures and features in imagery obtained by imaging sensors [27], [28]. Lastly, Boltzmann/Ising models have also been proposed as a convenient tool for distributed configuration management of sensors [29].

Apart from their great versatility, MRFs are also a desirable choice from an analytical point of view. This is because it is known, thanks to the classical Hammersley-Clifford theorem [30], that every MRF has an equivalent representation based on an exponential family [30]. In fact, exponential families can also describe the class of bi-partite graphical models called Factor Graphs [31], and much is known about their information-geometric properties [32]. Hence, without loss of generality, we will exclusively use the exponential family representation in formulating and solving the distributed identification problem.

B. Preliminary Setup

In the context of distributed detection [25], [33], it is often postulated that there is a single hidden hypothesis presented by nature and we have a large number of conditionally independent observations at spatially dispersed sites. We will consider a more general model where nature presents a spatial random process, and we make conditionally independent observations. Let $\{X^t_s\}$ be the hidden field drawn from an alphabet $\mathcal{X}$, and $\{Y^t_s\}$ be a related process observed by the sensors. Here the superscript $s \in \{1, 2, \ldots, N\}$ is the spatial index, indicating one out of $N$ spatially scattered sensors, while $t = 1, 2, 3 \ldots$ is a temporal index. For future reference let us define $X^t = [X^t_1, X^t_2, \ldots, X^t_N]^T$, with an analogous definitions for
While the observation process can be any kind of probabilistic measurement $Y^t_s = f(X^t_s, \omega)$ such that the conditional independence property $(Y^t_s \perp (X^t_{s'}, Y^t_{s'})) | X^t_s \forall (s', t') \neq (s, t)$ holds, for the sake of concreteness let

$$Y^t_s = X^t_s + V^t_s, \quad \forall s, t,$$

where $\{V^t_s\}$ is an additive Gaussian clutter process statistically independent of the field $\{X^t_s\}$, whose samples in space and time are independently and identically distributed with zero mean and variance $\sigma^2_V$.

Therefore $P(y^t|x^t) = \mathcal{N}(y^t|x^t, \sigma^2_V I_{N \times N})$, and $\mathbb{E}[|X^t_s|^2] / \sigma^2_V$ is the Signal to Clutter Ratio (SCR).

C. A Parametric Field Model

Suppose that $X^t, t = 0, 1, 2, \ldots$ is independently and identically distributed according to an MRF [5], where the direct statistical interaction of any spatial location with the rest of the field is limited to a physically localized neighborhood of that location. Conditioned on this statistical neighborhood, the field value at a location becomes independent of the rest of the field. Every such MRF implies an exponential family of p.m.f.s/ p.d.f.s [14]–[16], [30]

$$Q(x^t|\gamma) = \exp \left\{ \gamma^T b(x^t) - \Psi(\gamma) \right\},$$

where $\gamma \in \mathbb{R}^M$ is a parameter and $b : \mathcal{X}^N \rightarrow \mathbb{R}^M$ is a set of $M$ affinely independent basis functions. $\Psi(\gamma)$ is a normalization constant (the log-partition function). The Markovian property implies that each $b_i(x^t), i = 1, 2, \ldots, M$, depends only on a physically localized cluster of components from $x^t$. This set of locations is called the statistical neighborhood [5, Section II] of $i$. This is an extremely expressive model because every positive distribution of a discrete field $X^t$ can be accommodated, as well as many important continuous fields like GMRFs.

We will assume that the field satisfies three special constraints, namely: (a) Sparsity: The model has only $M = O(N)$ basis functions. (b) Localization: The statistical neighborhood of each location lies within a small ‘radius of interaction’ invariant w.r.t. $N$. (c) Mild Interactions: $\|\gamma\|_\infty \leq \Gamma_{\text{max}}$, where $\Gamma_{\text{max}}$ is some small number. Assumption ‘(a)’ is actually a default assumption in literature [5], [15], [16], made to avoid the ‘curse of dimensionality’. Assumption ‘(b)’ ensures that the communication load on each mote remains bounded even as the network scales, a prerequisite for scalability. It remains valid provided the network scales by increasing the spatial area of mote deployment while keeping the spatial density of motes roughly constant. Assumption ‘(c)’ allows MCMC methods like Gibbs Sampling to be terminated early, and is again a default assumption in literature [14]–[16]. Fortunately, these restrictions are far from artificial. Many natural phenomena of interest like diffusions and epidemics are indeed governed by sparse
Localized interactions are typically the rule rather than the exception, because the power of physical interactions (electromagnetic, acoustic or bio-chemical) falls off rapidly with distance. Finally, mild interactions are also common because un-modeled ‘nuisance’ processes typically preclude strong couplings.

D. An Example: The Boltzmann Field

To make matters concrete, consider the example of the Boltzmann field [34], where \( X = \{+1, -1\} \), and

\[
Q(x^t|\theta, W) = \exp \left\{ x^T \theta + \frac{1}{2} x^T W x^t - \Psi(\theta, W) \right\}.
\]

(3)

Here \( \theta \in \mathbb{R}^N \), and \( W \in \mathbb{R}^{N \times N} \) is a symmetric matrix with zeros along the diagonal. To see the correspondence with equation (2), note that \( \gamma \equiv (\theta, W) \), the first \( N \) elements of \( \gamma \) are \( \theta_1, \ldots, \theta_N \), and the next \( M - N \) elements are the \( M - N \) non-zero entries in the upper-triangular portion of \( W \), listed in some convenient order. Similarly the first \( N \) components of \( b(x^t) \) are \( x_1, \ldots, x_N \), and the next \( M - N \) components are the corresponding second order terms of the form \( x_i^t x_j^t, i \neq j \). The parameterization is minimal by construction (\( M \leq 2^N - 1 \)). \( M = O(N) \Rightarrow W \) is sparse. Physical localization implies that \( W \) has a structure that mirrors the topology of the WSN; this will be clarified with a numerical example in Section V. The statistical neighborhood of \( s \) is the subset of sites \( \mathcal{N}_s \subseteq \{1, 2, \ldots, N\} \) such that \( j \in \mathcal{N}_s \) iff \( W_{s,j} \neq 0 \).

III. Incremental Parameter Estimation, Stability and Covariance Efficiency

Structural knowledge of the model, i.e. the knowledge of its basis function, is often available from physical considerations. However, parametric knowledge, i.e. the actual numerical value of \( \gamma \), is rarely available a-priori, and needs to be blindly estimated based on the observations. Various researchers [11]–[16] have proposed, with minor variations, a stochastic EM algorithm (SAEM) as a tractable method to achieve this goal in lieu of a full-fledged EM algorithm [10]. We will, in particular, use a ‘partial M-step’ variant [18], where we start with some initial estimate \( \gamma^t \), which is recursively updated as

\[
\gamma^{t+1} = \gamma^t + \epsilon A S_{\gamma^t}(Y^t).
\]

(4)

In this recursion, \( S_{\gamma^t}(Y^t) \) is the score of \( Y^t \), \( \epsilon > 0 \) is a user-defined step size, and \( A \) is a constant user-defined symmetric positive-definite pre-scaling matrix (hence this recursion is not doubly stochastic like [16]). Please note that the necessary mathematical notation pertaining to exponential families is relegated to Appendix A, which the reader may wish to consult at this point. The rationale behind the choice of \( \epsilon \) and \( A \) will be discussed in Section III-B.
A. Asymptotic Stability of the Expected Gradient System

Using the Averaged-ODE method [35]–[37] it is sometimes possible to provide guarantees of almost-sure convergence of recursion (4) assuming a time-invariant truth-model, a suitably decaying sequence of step sizes, and certain strict regularity properties like scale invariance and global asymptotic stability. While realistic models may not necessarily possess all these regularity properties, expected gradient analysis nevertheless gives useful information about the stability and convergence of such algorithms. The basic premise is that, with a small enough step size, the trajectory of the stochastic recursion gets coupled to the trajectory of a deterministic recursion driven by an ‘averaged’ increment, where the averaging is done over all the sources of stochasticity. Denote the stochastic and expected gradients respectively as

\[ f(\gamma_t, Y_t) = A S_{\gamma_t}(Y_t) \]

and

\[ \bar{f}(\gamma_t) = \mathbb{E}_{\pi(y_t|\gamma^*)} [f(\gamma_t, Y_t)] \]

, where note that the expectation is done under the truth model with \( \pi(\cdot) \) as defined in Appendix A. Then we have the following theorem, which appears to be original:

**Theorem 1:** There exists a step size \( \epsilon > 0 \) and a threshold clutter variance \( \sigma^2_{V,\text{thresh}} > 0 \) such that, when \( \sigma^2_V < \sigma^2_{V,\text{thresh}} \), \( \gamma^* \) is as an asymptotically stable (A.S.) fixed point of the expected gradient recursion

\[ \gamma_{t+1} = \gamma_t + \epsilon \bar{f}(\gamma_t) \quad (5) \]

**Proof:** Define \( g(\gamma) \doteq \gamma + \epsilon \bar{f}(\gamma) \). Since the expectation of the score under the truth model is always zero [38], it follows that \( g(\gamma^*) = \gamma^* + \epsilon A \mathbb{E}_{\pi(y_t|\gamma^*)} [S_{\gamma^*}(Y_t)] = \gamma^* + 0M \). Hence \( \gamma^* \) is a fixed point of the recursion (5). To see that the recursion is asymptotically stable, note that

\[ \nabla_\gamma g(\gamma) \bigg|_{\gamma = \gamma^*} = I - \epsilon AF_{\gamma^*} \gamma . \]

Due to the property \( \lim_{\sigma_V \to 0} F^Y_{\gamma} = F^X_{\gamma} \) (cf. Appendix A) and the continuous dependence of the eigenvalues of a matrix on its elements, there exists a \( \sigma^2_{V,\text{thresh}} > 0 \) such that for \( \sigma^2_V < \sigma^2_{V,\text{thresh}} \) we have \( F^Y_{\gamma^*} > 0 \), and hence \( AF_{\gamma^*} > 0 \) (since \( A > 0 \) by hypothesis). Then, by Weyl’s theorem [39], there exists \( \epsilon > 0 \) such that all eigenvalues of \( \nabla_\gamma g(\gamma) \big|_{\gamma = \gamma^*} \) are inside the unit circle, which proves that \( \gamma^* \) is A.S. □

**Remark 1:** While the theorem above states that the expected gradient system is A.S. for a sufficiently large SCR, we conjecture, based on numerical calculations, that \( F^Y_{\gamma} > 0 \) and A.S. holds at all SCRs.

**Remark 2:** While \( l_\gamma(\cdot) \) is strictly concave in \( \gamma \), \( L_\gamma(\cdot) \) is not, since numerical evaluation of \( \frac{\partial \bar{f}(\gamma)}{\partial \gamma} \) at points sufficiently far away from \( \gamma^* \) yields a sign-indeterminate matrix. Thus \( \gamma^* \), in general, may not be globally A.S.
B. Efficiency Analysis

Let us define the error w.r.t. the averaged gradient recursion to be \( \alpha^t = \gamma^t - \bar{\gamma}^t \). Then by subtracting equation (5) from equation (4), we obtain the error evolution equation

\[
\alpha^{t+1} = \alpha^t + \varepsilon A \left( S_{\gamma^t}^t(Y^t) - \mathbb{E}_{\pi(y'|\gamma^t)} \left[ S_{\gamma^t}^t(Y^t) \right] \right).
\]

Recall that we showed in Theorem 1 that \( \bar{\gamma}^t \to \gamma^* \) provided \( \gamma^0 \) is in the attractor of \( \gamma^* \). We would like to analyze the error evolution in the regime of small deflections of \( \gamma^t \) around the putative lock point \( \gamma^* \).

Hence we assume that the error asymptotically converges to a wide-sense stationary phase with zero mean, and linearize equation (7) around zero error. That is, we expand \( S_{\gamma^*}^t(Y^t) \) and \( \mathbb{E}_{\pi(y'|\gamma^*)} \left[ S_{\gamma^*}^t(Y^t) \right] \) in a Taylor series around \( \gamma^* \) and then ignore second and higher order terms to get

\[
\alpha^{t+1} = (I - \varepsilon AF_{\gamma^*}^Y)\alpha^t + \varepsilon AS_{\gamma^*}^t(Y^t).
\]

Now we take second order moments on both sides and use the independence property \( \alpha^t \perp Y^t \), to obtain

\[
\mathbb{E} \left[ \alpha^{t+1}\alpha^{t+1T} \right] = (I - \varepsilon AF_{\gamma^*}^Y)\mathbb{E} \left[ \alpha^t\alpha^T \right] (I - \varepsilon AF_{\gamma^*}^Y) + \varepsilon^2 AF_{\gamma^*}^Y A.
\]

Let the time-invariant covariance matrix of the error in the stationary phase be denoted by \( \Sigma_\alpha \). Then it must satisfy the equation \( \Sigma_\alpha = (I - \varepsilon AF_{\gamma^*}^Y)\Sigma_\alpha(I - \varepsilon AF_{\gamma^*}^Y) + \varepsilon^2 AF_{\gamma^*}^Y A \), whose solution is given by

\[
\Sigma_\alpha = \varepsilon^2 \sum_{i=0}^{\infty} (I - \varepsilon AF_{\gamma^*}^Y)^i AF_{\gamma^*}^Y A (I - \varepsilon AF_{\gamma^*}^Y)^i.
\]

An important special case arises when we choose \( A = (F_{\gamma^*}^Y)^{-1} \), where the solution simplifies to

\[
\Sigma_\alpha = \frac{\varepsilon}{2 - \varepsilon}(F_{\gamma^*}^Y)^{-1}.
\]

Lemma 1: With \( A = (F_{\gamma^*}^Y)^{-1} \), the incremental estimator (4) is asymptotically efficient. □

Proof: Note that \( \frac{\varepsilon}{2 - \varepsilon} \) is the normalized noise equivalent bandwidth, and \( \tau(\varepsilon) = \frac{2 - \varepsilon}{\varepsilon} \) is the time constant, of a first order Infinite Impulse Response (IIR) filter \( x^t+1 = (1 - \varepsilon)x^t + \varepsilon u^t \). Since the incremental estimator in equation (4) is unbiased, the CRLB is applicable. However, for a fair comparison, one must postulate a data window \( n = \tau(\varepsilon) \). Now by choosing \( n \gg 1 \) (hence \( \varepsilon \ll 1 \)), we can ensure that linearized analysis holds with arbitrary precision. Then a comparison of equation (11) and the CRLB (20) makes it clear that the incremental estimator is asymptotically \( (n \to \infty, \varepsilon \to 0) \) efficient. □

Remark 3: To achieve the CRLB we need to use the ideal pre-scaling matrix \( A = (F_{\gamma^*}^Y)^{-1} \). This also has the useful consequence of making recursion (4) a “natural gradient” recursion [40], where the spread in the speeds of slow and fast manifolds is small, and convergence is uniform.

\[\text{January 16, 2009 DRAFT}\]
Remark 4: In practice, of course, we must presume some nominal value $\gamma^{nom}$ and set $A = (F_{\gamma^{nom}}^Y)^{-1}$. It is worth reemphasizing that the choice of a suboptimal $A > 0$ does not cause any loss in stability or unbiasedness (cf. Theorem 1). Hence we may intermittently recalculate $A = (F_{\gamma^{nom}}^Y)^{-1}$ where $\gamma^{nom}$ is the temporal average of the recent record of $\gamma^t$, and thus assure an efficiency close to the CRLB.

Remark 5: Even if we have a fairly good nominal estimate $\gamma^{nom} \approx \gamma^*$, we may still choose to use a sub-optimal pre-scaling matrix $A$ in order to satisfy certain localization constraints on message passing (c.f. Section IV). In this case, we must ensure the normalization $\| A \| = \| (F_{\gamma^{nom}}^Y)^{-1} \|$ is maintained, so that the estimator’s tracking speed is consistently determined by $\epsilon$ alone.

Remark 6: Numerical calculations indicate that when $\| \gamma^* \|$ is large the condition number of $F_{\gamma}^X = F_{\gamma}$ degrades. In light of the stability and efficiency analysis presented earlier, this implies that the relative stochastic stability of the recursion in equation (4) degrades. However note that the Kullback-Leibler (KL) divergence$^2$ of the estimated model w.r.t. the truth model is given by

$$D(\gamma^* \| \gamma^t) \equiv D(Q(x|\gamma^*) \| Q(x|\gamma^t)) = \frac{1}{2} (\gamma^* - \gamma^t)^T F_{\gamma^*}^X (\gamma^* - \gamma^t) + \text{H.O.T. [nats]}. \quad (12)$$

It is known [38] that the KL-divergence, rather than the mean squared error (MSE), is the key information theoretic measure for quantifying the mismatch loss in hypothesis testing. Hence, depending on $F_{\gamma}^X$, a large MSE need not necessarily translate into a proportionately large KL-divergence. Secondly, though a small $F_{\gamma}^X$ leads to a large MSE via the CRLB, its presence in equation (12) mitigates the final effect on $D(\gamma^* \| \gamma^t)$.

IV. DISTRIBUTED IMPLEMENTATION, SCALABILITY AND POWER EFFICIENCY

In Section III we showed that the proposed estimator has desirable properties of stability and covariance efficiency. Now we will tackle the question of distributed implementation and power efficiency.

A. Choice of Algorithm For Calculating Expectations

The calculation of the gradient, $A S_{\gamma^t}(Y^t)$, is, in general, an NP-hard problem w.r.t. $N$, and the use of some kind of approximation is mandatory. Moreover, we should exploit the broadcast communication characteristic of the WSN. Four kinds of broadcast message-passing approximate marginalization algorithms are known in literature (see [7] and references therein): Gibbs Sampling (GS), Mean Field Decoding (MFD), Iterated Conditional Models (ICM) and Broadcast Belief Propagation (BBP). Unfortunately, while BBP,

$^2$The KL divergence of a distribution $f_2$ relative to a distribution $f_1$ is defined as $D(f_1(x) \| f_2(x)) \equiv \mathbb{E}_{f_1(x)} \left[ \log \frac{f_2(x)}{f_1(x)} \right]$. 

January 16, 2009 DRAFT
MFD and ICM have low complexity and give good performance for distributed filtering, they are not well suited to parameter estimation because they produce biased estimates of expectations of higher order (non-linear) basis functions, which leads to biases and instabilities in the recursive estimator (4). GS on the other hand always gives stable and unbiased estimates for the expectations of all the basis functions. Hence, similar to [13] [14] [15] [16], we prefer its use for calculating the gradient needed by the recursion. In Section IV-B we will describe the distributed implementation of the incremental estimator based on GS, and in Section IV-C we will analyze its total power consumption in detail.

B. Distributed Scalable Computation of the Gradient Using GS

For clarity of exposition, we will describe the gradient computation for the specific example of the Boltzmann model from Section II-D where \( \gamma^t \equiv (\theta^t, W^t) \). It is easy to generalize, by analogy, to other exponential models. It is known that GS is a stochastic method that yields an arbitrarily good approximation of the expectation of any function of a random variable \( X \) drawn according to \( Q(x|\gamma^t) \). For calculating this approximation, sufficiently many transitions are made of a specially constructed Markov Chain (MC) \( \{\zeta^k\} \). We will say that one iteration has been completed when \( N \) transitions of the MC are made, and let \( n_{\text{iters}} \) denote the total number of such iterations. The state space of \( \{\zeta^k\} \) is \( \{+1, -1\}^N \). In the simplest construction, from the state \( \zeta^k \) at the \( k \)-th iteration, outgoing transitions are allowed only by flipping at most one component of \( \zeta^k \). The probability of making a transition by flipping the \( i \)-th component is given in terms of the ‘full conditionals’ derived from \( Q(x|\gamma^t) \),

\[
\Pr\{[a_1, \ldots, a_i, \ldots, a_N]^T \to [\bar{a}_1, \ldots, \bar{a}_i, \ldots, a_N]^T\} = c \exp\left( \bar{a}_i [\theta^t_i + \sum_{j \neq i} W^t_{ij} a_j] \right),
\]

where \( c \) is a normalizing constant. The MCMC has the invariant distribution \( Q(x|\gamma^t) \), and the empirical average of \( b(\zeta^k) \) converges a.s. to its expectation \( \eta_{\gamma^t} \) [17]. We can thereby calculate the conditional and unconditional means, \( \eta_{\gamma^t+h(Y^t)} \) and \( \eta_{\gamma^t} \), and the score \( S_{\gamma}(Y^t) \), according to equations (16) and (19).

Consider first the calculation of \( \eta_{\gamma^t} \). A sensor node with spatial index \( s \) merely needs to maintain the MCMC state component \( \zeta_s \), which corresponds to the field variable \( X^t_s \). In one iteration, each site \( s \) receives the values of all the other local state variables from its statistical neighborhood. Each site \( s \) then computes a new realization for \( \zeta_s \), according to the distribution in equation (13), and broadcasts it back to all the sites in its statistical neighborhood. The order of the site updates is not critical, and may be chosen pseudo-randomly. After sufficiently many \( (n_{\text{iter}}) \) such iterations, the temporal averages are read-off as approximations of the true expectations of the respective basis functions. Similarly, \( \eta_{\gamma^t+h(Y^t)} \) can be calculated by running a GS that re-samples sites based on the distribution \( Q(x|\gamma^t+h(Y^t)) \).
This shows that we can make a distributed scalable calculation of the per-sample score \( S_{\gamma t}(Y^t) = \eta_{\gamma t} + h(Y^t) - \eta_{\gamma t} \). For the time being suppose \( A = I \). Since we can maintain the components of the estimated parameter vector \( \gamma^t \) at respective local sites, we then already have a fully distributed (local, component-wise) implementation of the recursion in equation (4). Now suppose we use a non-trivial \( A \), hence a “natural gradient” algorithm. We can still implement the recursion in equation (4) in an efficient distributed fashion provided we ensure that the non-zero entries in \( A \) have a structure such that communication is needed only between sensors in close physical proximity, say one or two hops apart. Thus note that there are two conflicting demands in choosing \( A \): On one hand, to achieve covariance efficiency (in the Cramér-Rao sense), it needs to be chosen close \((F_{\gamma}^X)^{-1}\). At the same time, to achieve power efficiency, it needs to be sparse in a particular way. Fortunately, \((F_{\gamma}^Y)^{-1}\) is already almost sparse in the required sense provided \( W \) is sparse with localized interactions (cf. model assumptions in Section II-C) and the SCR is not too small. A very instructive special case arises for a Gauss-Markov random field [23], where this statement is exactly true because \((F_{\gamma}^X)^{-1} = -W\), where \( W \) is now called the sensitivity matrix.

C. Power Efficiency

The total power consumption of the distributed implementation of GS is determined by three factors, namely, the frequency of the estimator updates, the number of GS iterations per update, and the communication load per iteration per update. We will now analyze the total power consumption and make comparisons with a centralized estimator.

1) Frequency of Estimator Updates: The statistical model of a natural field is typically quasi-static, i.e. it changes with a time constant of \( \chi \) sampling intervals, where \( \chi \gg 1 \). For adequate tracking we only need to ensure that \( \tau(\epsilon) \leq \chi \), hence \( \epsilon \geq \frac{\chi/\epsilon}{\chi + 1} \). However, the same real-time tracking performance (in seconds) can also be achieved if we choose any \( n_{\text{update}} \geq 1 \), set \( \epsilon \geq \frac{2n_{\text{update}}}{\chi + 1} \), and update the parameter only once every \( n_{\text{update}} \) samples. By using an \( n_{\text{update}} > 1 \) we can improve power efficiency by a factor of \( n_{\text{update}} \), while sacrificing covariance efficiency by the same factor, a technique we call interlacing.

2) Number of GS Iterations Per Update: It is known that, depending on how strongly the MC ‘mixes’, the number of GS iterations, \( n_{\text{iters}} \), may need to be in the thousands to produce very accurate expectations. However in practice we can make an early termination of the GS iterations in each update, paying with an increased covariance of the calculated per-sample score, but still maintaining unbiasedness.

3) Communication Load Per Iteration Per Update: The statistical neighborhood of each mote mirrors a small physical neighborhood (cf. Section II-C). Owing to the nature of wireless communication, the “communication graph” is also determined by physical proximity. Hence the statistical neighborhood can
be covered by a single broadcast transmission or perhaps a small number of hops, \( n_{\text{hops}} \) [8, page 64]. Thus the communication load per mote per iteration per update remains invariant as \( N \) increases.

4) Analysis of Power Budget With Multi-Hop Message-Passing: Let the radius of interaction in the field be \( r \) meters. Let the sampling interval be \( T_{\text{samp}} \) seconds. Assume that the transmission of a message from a mote to another mote is done over an additive white Gaussian noise (AWGN) channel and allowed a time slot of \( T \) seconds and a bandwidth of \( W \) Hz. Thus one mote transmission consists of \( n_S = WT \) quadrature symbols. In the scenario of multi-hop communication between motes with \( n_{\text{hops}} \geq 1 \) hops, we will pessimistically assume that an end-to-end transmission must be completed in \( T \) seconds, implying that each hop of \( r/n_{\text{hops}} \) meters is allowed only \( T/n_{\text{hops}} \) seconds. Let \( \varrho \) denote the path-loss exponent for all the wireless transmissions [41]. Denote the noise power spectral density (PSD) of the front-end receiver in a mote by \( N_0 \) Watts/Hz, and define \( \sigma_{\text{noise}}^2 = TWN_0 \). Lastly, let \( P_R \) be the power consumed by the receiver in each mote under the pessimistic assumption that it is always ‘ON’. \( P_R \) is obviously independent of \( N \).

In order to implement recursion (4) with a distributed Gibbs Sampler, each mote needs to communicate \( 2n_{\text{iters}} \) one-bit messages over a maximal distance of \( r \) meters. (One half each for the Conditioned-GS and the Unconditioned-GS). Note that the communication of the GS messages is delay constrained, hence we must assume that the messages are not compressed. Similarly, long channel codes cannot be used, due to which an energy penalty \( \lambda \) is suffered w.r.t. Shannon capacity. Using the same technique as in [7], which was based on Shannon’s capacity theorem [38] and the Friis path-loss model [41], we can derive the following upper bound on power consumption of each mote while implementing the incremental estimator of equation (4) in a distributed fashion:

\[
P_{\text{distrib}} \leq P_R + 2 \frac{\lambda n_{\text{iters}}}{n_{\text{update}} T_{\text{samp}} \sigma_{\text{noise}}^2 r^\varrho} \frac{2^{n_{\text{hops}}} - 1}{n_{\text{hops}}^{\varrho-1}} \quad [\text{Watts}].
\]

Note that \( n_{\text{hops}} \) is a constant independent of \( N \), as has been already discussed in 3) above. Similarly, we will see simulation evidence in Section V-C that even \( n_{\text{iters}} \) can be chosen to be a constant independent

### TABLE I

(A) Upper bound on increase in power consumption due to multi-hop message passing. (B) Lower bound on system gain of distributed Vs centralized model estimation.

<table>
<thead>
<tr>
<th>( k )</th>
<th>( n_S = 4 )</th>
<th>( n_S = 8 )</th>
<th>( n_S = 16 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.09</td>
<td>1.04</td>
<td>1.02</td>
</tr>
<tr>
<td>3</td>
<td>1.20</td>
<td>1.09</td>
<td>1.04</td>
</tr>
<tr>
<td>4</td>
<td>1.31</td>
<td>1.14</td>
<td>1.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n_{\text{iters}} )</th>
<th>( d ) = 100</th>
<th>( d ) = 1000</th>
<th>( d ) = 10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>18.7</td>
<td>38.7</td>
<td>58.7</td>
</tr>
<tr>
<td>100</td>
<td>15.7</td>
<td>35.7</td>
<td>55.7</td>
</tr>
<tr>
<td>200</td>
<td>12.7</td>
<td>32.7</td>
<td>52.7</td>
</tr>
</tbody>
</table>
of $N$. Hence it follows that bound (14) is independent of the network size $N$, thus demonstrating that our algorithm is strictly scalable. Secondly, the power consumption varies as $\frac{n_{\text{hops}}}{n_{\text{update}}}$, and this allows us to choose any desirable tradeoff w.r.t. Cramer-Rao efficiency. Thirdly, the bound also demonstrates that if the statistical neighborhood needs to be covered by $n_{\text{hops}} > 1$ hops, then in the worst case the power consumption is increased by a factor $\frac{2^{n_{\text{hops}}}-1}{(2^{n_{S}}-1)n_{\text{hops}}^{n_{\text{hops}}}}$, which is quite small when $n_{\text{hops}} = 1 \sim 4$. For example, Table I(A) demonstrates the increase in power consumption due to multi-hop communication for the case of $\varphi = 2$ (free space propagation) and several values of $n_{S}$. We conclude that the communication of GS messages with a small number of hops is not a fundamental drawback.

5) Distributed vs Centralized Estimation: We will now analyze the system gain that the distributed identification algorithm enjoys over a centralized algorithm. Note that the details of the centralized algorithm are immaterial to the power efficiency comparison because the energy cost incurred in centralized processing is solely attributed to the aggregation of raw data at the FC; there is no ‘algorithmic’ energy cost.

Analogous to the assumptions made in 4) above, let $\hat{N}_0$ be the noise PSD of the front-end receiver in the FC, and let $\hat{\sigma}_{\text{noise}}^2 = TW \hat{N}_0$. Let the distance of the WSN to the FC be $d$ meters. We will assume that the communication with the FC is done by direct long distance transmissions, without repeaters. This is a valid assumption when the FC is far removed from the WSN array ($d \gg r$). In a centralized scheme, after every $n_{\text{update}}$ sampling intervals, we need to extract all the sensor data to the FC. This implies transmitting $N$ real numbers, quantized to some sufficient precision $\frac{1}{2^{n_{\text{acc}}}}$, over a distance of $d$ meters. Let us assume (optimistically, from the point of view of the centralized scheme) that the data is optimally compressed with Slepian-Wolf type non-cooperative encoders, and a capacity achieving channel code is used. We will also assume that a suitable time sharing schedule is used so that the burden of long distance transmissions is distributed evenly among the motes. Let $P_{\text{central}}$ denote the power spent by each mote on average in the centralized estimator implementation. Then we can loosely lower bound the “system gain” as

$$\frac{P_{\text{central}}}{P_{\text{distrib}}} \geq \frac{d^{\varphi} \hat{\sigma}_{\text{noise}}^2}{n_{\text{update}}T_{\text{samp}}P_{R} + \frac{2n_{\text{hops}}}{n_{S}}r^{\varphi} \hat{\sigma}_{\text{noise}}^2 \frac{(2^{n_{\text{hops}}}-1)}{(2^{n_{S}}-1)n_{\text{hops}}}}$$

(15)

Since $r$ and $P_{R}$ are constants, as $d \to \infty$ the gain increases monotonically and unboundedly implying that there is always a break-even distance where it exceeds unity. Table I(B) gives the power gain for exemplary values of $n_{\text{iters}}$ and $\varphi$, and a set of practical values for the other parameters as follows: $\varphi = 2.0, \hat{\sigma}_{\text{noise}}^2 = \hat{\sigma}_{\text{noise}}^2, \lambda = 10, n_{\text{acc}} = 4, n_{S} = 4, n_{\text{hops}} = 1, P_{R} = 0$ (negligible receiver consumption) and $\sigma^2_{V} = -6$ dB. Clearly, on the basis of power efficiency, there is a strong incentive for in-network model estimation.
V. Simulation Results and Discussion

In this section we present comprehensive simulations for an exemplary application where a uniform linear array of $N$ sensors measures the presence (+1) or absence (−1) of an effluent released from a chemical plant into a river. We use a Boltzmann field model (cf. Section II-D) for the statistical distribution of the effluent, where the spatial dependencies are given by an $N \times N$ Toeplitz matrix parameter whose first row is $\rho[0, \xi_1, \xi_2, \xi_3, 0, \ldots, 0]$. We assume that $\theta = 0_N$, implying that the presence or absence at any location is a-priori equiprobable. The statistical neighborhood of the above model is 3 units. The parameter $\rho > 0$ is used to control the strength of the spatial interactions, with a large $\rho$ leading to large plumes on average. Moderate dependencies refers to the choice $\rho = 1.0$, while strong dependencies refers to $\rho = 2.0$. Note that we postulate a spatial homogeneous interaction (i.e. a Toeplitz structure) only for simplicity of exposition. The algorithm itself does not presume such a property; indeed, each entry of $W$ can in principle be distinct. The total truth parameter is $\gamma^* \equiv (\theta^*, W^*)$, of length $M = 4N - 6$. Static model will refer to the choice $[\xi_1, \xi_2, \xi_3] = [0.5, 0.3, -0.3]$, while Time-varying model of period $\chi$ will refer to the choice $[\xi_1, \xi_2, \xi_3] = [0.5, 0.3, -0.3] \times \left(1 + \frac{1}{2} \sin \frac{2\pi t}{\chi}\right)$. In the latter case, $\gamma^{nom} \equiv [0.5, 0.3, -0.3]$.

A. Covariance Efficiency

First we will study the steady state variance-covariance efficiency vs $\tau(\epsilon)$ characteristic, as a function of the SCR, the update interval $n_{update}$, the number of GS iterations $n_{iters}$, and the type of constraints on the pre-scaling matrix $A$. As nominal values we choose SCR = 3 dB, $n_{iters} = 128$, $n_{update} = 1$ and $A = (F_{\gamma^*}^Y)^{-1}$ with no constraints. We vary each parameter individually while keeping the others fixed at their nominal values, and the results are presented in Figure 1(a)-(d). We also plot the CRLB as a baseline reference in each case. The first column in each sub-figure corresponds to moderate dependencies and the second column to strong dependencies.

We observe that (a) In Figure 1(a), for all practical SCRs and tracking speeds there is very good agreement between linearized analysis (i.e. the CRLB) and simulations. The simulations degrade from the CRLB for large $\epsilon$ because the estimator starts making large excursions, and sometimes loses ‘lock’. It is interesting to note that the reduction in efficiency due to a lowering of the SCR is not dB-to-dB. Also, as an example, for the moderate dependencies case we have demonstrated that the MSE of the sub-components, i.e. $\frac{1}{N}\mathbb{E}||\theta^t - \theta^*||^2$ and $\frac{1}{M-N}\mathbb{E}||W^t - W^*||^2_2$ labelled as “$\theta$ only” and “$W$ only” respectively, is qualitatively similar to the total MSE. This is a consequence of the optimal pre-scaling operation. (b) As expected, in Figure 1(b), interlacing with a factor $n_{update}$ causes the entire performance characteristic to shift right by an equal factor. (c) As predicted in Section IV-B, in Figure 1(c) the covariance efficiency is seen to approach
the CRLB when $n_{\text{iters}}$ is sufficiently large. The loss from the CRLB due to an insufficient $n_{\text{iters}}$ depends on the tracking time-constant and the strength of the field dependencies. Interestingly, we find that even with $\tau(\epsilon) \approx 20000$, and a strong field dependencies, a practical choice like $n_{\text{iters}} = 128$ gives a small loss of $1.0$ dB w.r.t. the CRLB. (d) Finally, in Figure 1(d) we postulate a user defined locale $l$ (in units of distance), and obtain the pre-scaling matrix $A$ by masking $(F_\gamma^Y)^{-1}$ with a localization matrix that allows communication only between sensors that are within $l$ units of each other. Thus, $l = 3.0$ implies that a mote is allowed to communicate with another mote no more than 3.0 units apart, while $l = \infty \Rightarrow A = (F_\gamma^Y)^{-1}$ exactly. We observe that the use of a locale equal to the radius of interaction causes very minor loss in performance. Using a locale of $l = 0$, hence a diagonal pre-scaling matrix, causes a more significant loss in efficiency, of the order of $4.0$ dB. Thus we see that using an $A$ that conforms with $W$ gives most of the available efficiency, without putting any extra communication load on the motes (cf. Section IV-B).

B. Model Acquisition and Tracking

Now we investigate dynamic performance of the estimator, under a time varying model with a time constant $\chi = 16000$ and moderate dependencies. We let $A = (F_\gamma^Y_{\text{nom}})^{-1}$, $\epsilon = 10^{-5}$, $\text{SCR} = 9.0$ dB. In Figure 2, in the first sub-plot we show the time varying truth parameter (bold lines), and the estimated parameter (thin lines). In the second sub-plot, we show the KL-divergence between the truth model and the estimated model, $D(\gamma^*\|\gamma^t)$, and also between the truth model and the nominal model $D(\gamma^*\|\gamma^\text{nom})$. The simulations are started from the maximally uninformative initial condition $\gamma^1 = 0$, so that we can also observe the acquisition dynamics (the step response). We observe that (i) The acquisition time agrees very well with the estimator time constant $\tau(\epsilon) = \frac{2-\epsilon}{\epsilon} = 2000$, and the estimator can consistently track the model because $\chi > \tau(\epsilon)$. (iii) An uncompensated (open-loop) system has large excursions in KL-divergence (upto 0.35 bits), while the adaptive estimator gives comparatively small excursions (upto 0.05 bits). Similarly, the MSE of the uncompensated case can rise upto $-15$ dB per component, while the MSE of the estimator never rises above $-23$ dB. (iv) There are no pathologies like ringing or under-damping.

C. Scalability

Finally we will consider the important issue of scalability. We have already seen that even if the network size $N$ is increased, the computation and communication load on each mote remains invariant. Here we investigate whether the scalability also holds in terms of the performance. We choose a static truth model with moderate dependencies and essentially repeat the experiment in Section V-B with several values of $N$. Each experiment is started from the initial condition $\gamma^1 = 0$. Since it is difficult to exactly calculate $F_\gamma^Y$
for large networks (say \( N \geq 32 \)) we will use a suboptimal diagonal pre-scaling matrix \( A \). The values on the main diagonal of \((P^Y_\gamma)^{-1}\) for \( N = 8 \) are interpreted as weights for the corresponding basis functions, and in constructing \( A \) for a larger \( N \) these weights are appropriately replicated for all the corresponding extra basis functions in the model. The spatial homogeneity of our model ensures that such a construction of \( A \) gives a fair comparison between networks of various sizes. We set \( \epsilon = 10^{-3} \). In Figure 3, we show the normalized MSE of the estimator as a function of time, for various \( N \). Subplot (a) is for \( n_{\text{iters}} = 128 \) and (b) is for \( n_{\text{iters}} = 8 \). We make the crucial observation that the acquisition dynamics and the steady state value of the MSE is essentially invariant w.r.t. \( N \), even for a very small number of MCMC iterations \( n_{\text{iters}} = 8 \). The increase in the MSE from \( n_{\text{iters}} = 128 \) to \( n_{\text{iters}} = 8 \) is small (about 2.0 dB) and is the same for all network sizes. Therefore, the estimator acquires and tracks in an identical manner, irrespective of the network size, provided \( n_{\text{iters}} \) is a little larger than the statistical neighborhood in the model.

VI. Conclusions

We have proposed a distributed incremental estimator for exponential models in Wireless Sensor Networks. We have shown that the algorithm is stable, asymptotically efficient, strictly scalable, and has a power consumption significantly smaller than a centralized estimator even in the scenario of multi-hop message passing. Note that it is possible to extend our model identification scheme in a straightforward manner to the case of fields with spatial as well as temporal dependencies [42].

APPENDIX A

EXPONENTIAL FAMILIES: SOME NOTATION AND PROPERTIES

Let \( X \sim Q(x|\gamma) \), where \( Q(\cdot|\gamma) \) is as defined in equation (2), and let \( Y = X + V \) where \( V \sim N(0_N, \sigma^2_V I_{N \times N}) \). Let \( \pi(y|\gamma) = \sum_x Q(x|\gamma)P(y|x) \) denote the unconditional distribution of the observation \( Y \) parameterized by \( \gamma \). Let \( \eta_\gamma = \nabla_\gamma \Psi(\gamma) \) and \( F_\gamma = \nabla^2_\gamma \Psi(\gamma) \). It is known [32] that \( \eta_\gamma \) and \( F_\gamma \) are respectively the mean and covariance in the directly observed model,

\[
\eta_\gamma = \mathbb{E}_{Q(x|\gamma)}[b(X)], \quad F_\gamma = \mathbb{E}_{Q(x|\gamma)}[(b(X) - \eta_\gamma)(b(X) - \eta_\gamma)^T].
\]  

The direct and indirect observation log-likelihoods are \( l_\gamma(X) \doteq \log Q(X|\gamma) \) and \( L_\gamma(Y) \doteq \log \pi(Y|\gamma) \), the respective scores are \( s_\gamma(X) = \nabla_\gamma l_\gamma(X) \) and \( S_\gamma(Y) = \nabla_\gamma L_\gamma(Y) \), and the respective Fisher informations are \( F^X_\gamma = \mathbb{E}_{Q(x|\gamma)} [s_\gamma(X)s_\gamma(X)^T] = -\mathbb{E}_{Q(x|\gamma)} [\nabla^2_\gamma l_\gamma(X)] \) and \( F^Y_\gamma = \mathbb{E}_{\pi(y|\gamma)} [S_\gamma(Y)S_\gamma(Y)^T] = -\mathbb{E}_{\pi(y|\gamma)} [\nabla^2_\gamma L_\gamma(Y)] \) [32], [38], [43]. It is straightforward to show that \( F^X_\gamma = F^Y_\gamma \). Also note that when \( \sigma^2_V = 0 \), we have direct observation \( Y = X \), and therefore \( S_\gamma(Y) = s_\gamma(X) \) and \( F^Y_\gamma = F^X_\gamma \). Since \( F^Y_\gamma \)
is a continuous function of $\sigma^2_\gamma$. we have $\lim_{\sigma^2_\gamma \to 0} F^Y_\gamma = F^X_\gamma$, where the convergence is element-wise. An important property of regular (i.e. minimal) exponential families is that $\Psi(\gamma)$ is strictly convex [43] in $\gamma$, and hence the log-likelihood $L_\gamma(\cdot)$ is strictly concave in $\gamma$ for any $X$. This means that $\forall \gamma, F^X_\gamma = F_\gamma > 0$. For Boltzmann fields ($\mathcal{X} = \{+1, -1\}$), we can define an indirect-observation log-likelihood ratio

$$h_i(Y) = \begin{cases} \frac{1}{2} \log \frac{P(Y_i|x_i = +1)}{P(Y_i|x_i = -1)} = \frac{Y_i}{\sigma^2_\gamma}, & i = 1, 2, \ldots, N \\ 0_{M-N}, & i = N + 1, N + 2, \ldots, M. \end{cases} \tag{17}$$

Then it is easily shown that [32]

$$s_\gamma(X) = b(X) - \eta_\gamma \tag{18}$$
$$S_\gamma(Y) = E_{P(x|Y,\gamma)}[s_\gamma(X)] = \eta_{\gamma+h(Y)} - \eta_\gamma. \tag{19}$$

Finally, suppose that $Q(\cdot|\gamma^*)$ is the ‘truth model’. Then, if $\hat{Y}^n = [(Y^1)^T, (Y^2)^T, \ldots, (Y^n)^T]^T$, the Maximum Likelihood (ML) estimate of $\gamma^*$ is $\hat{\gamma}_{ML}(n) = \max_{\gamma \in \Gamma_{\text{feasible}}} \log P(\hat{Y}^n|\gamma)$. It is known [44] that $\hat{\gamma}_{ML}(n)$ is asymptotically (as $n \to \infty$) unbiased and efficient i.e. achieves the Cramér-Rao lower bound (CRLB)

$$E \left[ (\hat{\gamma}_{ML}(n) - \gamma^*)(\hat{\gamma}_{ML}(n) - \gamma^*)^T \right] \geq \frac{1}{n} (F^Y_{\gamma^*})^{-1}. \tag{20}$$

REFERENCES


17


Fig. 1. Dependence of the variance efficiency on (a) SCR (b) update interval, $n_{\text{update}}$ (c) number of GS iterations, $n_{\text{iters}}$ (d) constraints placed on the pre-scaling matrix $A$ ($l$ denotes the locale used to mask $(F_Y^*)^{-1}$). Nominal parameters: $N = 8$, $\text{SCR} = 3.0, n_{\text{update}} = 1, n_{\text{iters}} = 128, A = (F_Y^*)^{-1}$. 

January 16, 2009 DRAFT
Fig. 2. Acquisition and tracking performance under a time-varying model with period $\chi = 16000$ samples. $N = 8$, SCR = 9.0 dB, $n_{\text{update}} = 1$, $n_{\text{iters}} = 128$, $A = (P_{\gamma_{\text{com}}}^{-1})^t$.

Fig. 3. Scalability w.r.t. the size of the network $N$. $n_{\text{iters}} = 8$, SCR = 9.0 dB, $\epsilon = 10^{-3}$, $n_{\text{update}} = 1$ and a diagonal pre-scaling matrix $A$ (cf. Section V-C). Bold lines indicate $\gamma^*$, and thin lines indicate $\gamma^t$. (a) $n_{\text{iters}} = 128$ (b) $n_{\text{iters}} = 8$. 

January 16, 2009