AN EFFICIENT MONTE CARLO APPROACH FOR OPTIMIZING DECENTRALIZED ESTIMATION NETWORKS CONSTRAINED BY UNDIRECTED TOPOLOGIES

Murat Üney[†‡], *Müjdat Çetin*[†]

[†]Sabancı University, Faculty of Engineering and Natural Sciences,
Orhanlı-Tuzla, 34956, İstanbul, Turkey.
[‡]Middle East Technical University, Department of Electrical and Electronics Engineering,
06531, Ankara, Turkey.

ABSTRACT

We consider a decentralized estimation network subject to communication constraints such that nearby platforms can communicate with each other through low capacity links rendering an undirected graph. After transmitting symbols based on its measurement, each node outputs an estimate for the random variable it is associated with as a function of both the measurement and incoming messages from neighbors. We are concerned with the underlying design problem and handle it through a Bayesian risk that penalizes the cost of communications as well as estimation errors, and constraining the feasible set of communication and estimation rules local to each node by the undirected communication graph. We adopt an iterative solution previously proposed for decentralized detection networks which can be carried out in a message passing fashion under certain conditions. For the estimation case, the integral operators involved do not yield closed form solutions in general so we utilize Monte Carlo methods. We achieve an iterative algorithm which yields an approximation to an optimal decentralized estimation strategy in a person by person sense subject to such constraints. In an example, we present a quantification of the trade-off between the estimation accuracy and cost of communications using the proposed algorithm.

***Index Terms*—** Decentralized estimation, communication constrained inference, random-field estimation, message passing algorithms.

1. INTRODUCTION

Decentralized estimation underlies many envisioned applications of sensor networks which are networked platforms that have limited capability of sensing, communication and computation. Possible scenarios consider a relatively high volume of data collected at various locations often in an uncollaborating environment. Therefore, platforms need to communicate through bandwidth (BW) limited links in order to have the data processed. Besides, the limited energy budget is mostly consumed by the transmissions. Also the processing is preferred to be done in a collaborative fashion to inhibit possible computational bottlenecks and decrease BW requirements. Hence, the issues regarding the achievable estimation accuracy for a given communications structure and transmission costs together with the decentralized strategy that exhibits a certain performance arise.

The conventional setting renders a star shaped directed graph, in which a fusion center is selected to perform the estimation task depending on the quantized observations collected and transmitted by the peripheral nodes (see e.g. [1, 2, 3]). The design problem involves choosing the quantization schemes together with a fusion rule that exhibit a certain performance. Altough BW constraints are considered, the cost of transmissions which likely vary for each link due to the multi-hop nature and more general topologies which might better reflect an ad-hoc setting are not captured under these treatments. Also, in the case of multiple random variables, e.g. as in a random-field estimation problem, computational bottleneck problems might occur at the fusion center and the lack of collaboration among nodes might inhibit the improvement of the performance.

If the underlying network services support a relatively high load, Graphical Models together with Message Passing Algorithms provide solutions in accordance with the in-network processing paradigm [4]. Altough it is possible to analyze the effects of the communication structure in this framework [5], it not easy to tailor the solution given the communication constraints.

We consider the estimation of an N-dimensional random vector by a distributed system which exhibits a communication and computation structure that better matches the underlying ad-hoc, multi-hop nature. We are concerned with introducing the cost of communications, possibly due to energy consumption, as well as the availability and capacity of links. A collaborative processing is achieved through distributing the estimation task through random variable-node associations. A Bayesian approach in which the costs both due to communications and estimation errors are captured provides a rigorous problem definition. Such a setting is utilized in [6] for the case in which the underlying communications render a directed graph. In this work, we consider bidirectional links rendering an undirected graph (UG). For detection networks, a similar design problem has been investigated in [7] (see also [8]) in which rules local to nodes for communication as well as detection are sought such that a dual objective Bayesian risk is optimized. The aggregation of local rules are called a strategy and the set of feasible strategies is constrained by the UG structure. Under a Team Decision Theoretic treatment an iterative solution which converges to a person by person (pbp) optimal strategy is proposed. We adopt this framework for decentralized estimation (DE) and present the corresponding iterative scheme. However the resulting expressions contain integral operators which are impossible to evaluate exactly in practice. In order to keep fidelity to the mathematical model, we exploit Monte Carlo (MC) integration methods and achieve a MC optimization scheme for DE networks constrained by an UG which is scalable with the number of nodes and sample sizes. Moreover, results can be produced for any set of distributions provided that samples can be generated from them. The resulting strategy corresponds to approximate computations to the pbp optimal one achieving a reasonable Bayesian risk.

This work was partially supported by the Scientific and Technological Research Council of Turkey under grant 105E090, by the European Commission under grant MIRG-CT-2006-041919 and with a Turkish Academy of Sciences Young Scientist Award. The authors would like to thank O. Patrick Kreidl for his help and support throughout many discussions.

978-1-4244-2710-9/09/$25.00 © 2009 IEEE

2. THE DESIGN PROBLEM

We consider the estimation counterpart of the decentralized detection network design problem considered in [7]. Hence, in our setting the variables to be inferred take values from denumerable sets. We assume that the links are error-free.

2.1. Online Processing Constrained With an UG

Representing a set of platforms with the index set $\mathcal{V} = \{1, ..., N\}$, with each $j \in \mathcal{V}$ a random variable X_j is associated that takes values from the set $\mathcal{X}_j$ which, unlike the detection case, is denumerable. $X = (X_1, ..., X_N)$ is the random field of concern where a realization x satisfies $x \in \mathcal{X}$ with $\mathcal{X} = \mathcal{X}_1 \times ... \times \mathcal{X}_N$. Given a set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an UG if it holds that $(i, j) \in \mathcal{E}$ implies $(j, i) \in \mathcal{E}$. Given $\mathcal{G}$, each edge $(i, j) \in \mathcal{E}$ corresponds to a communication link of capacity $\log_2(|\mathcal{U}_{i\to j}| - 1)$ bits such that $\mathcal{U}_{i\to j}$ is the set of admissible symbols with the symbol $0 \in \mathcal{U}_{i\to j}$ indicating no transmission.

Let $u_{ne(j)} \triangleq \{u_{i\to j} | i \in ne(j)\}$ denote the incoming messages to node j from neighbor nodes $ne(j)$, which takes values from $\mathcal{U}_{ne(j)} = \mathcal{U}_{ne_j^1 \to j} \times ... \times \mathcal{U}_{ne_j^D \to j}$. Here $ne(j) = \{ne_j^1, ..., ne_j^D\}$. The outgoing messages from node j to neighbor nodes $ne(j)$ is given by $u_j \triangleq \{u_{j\to i} | i \in ne(j)\}$ and takes values from $\mathcal{U}_j$ which can be defined similarly with that for $\mathcal{U}_{ne(j)}$. The overall communication load is $u \triangleq \{u_{i\to j} | (i, j) \in \mathcal{E}\}$ and takes values from $\mathcal{U} = \mathcal{U}_1 \times ... \times \mathcal{U}_N$.

A causal online processing of measurements $\{y_j | j \in \mathcal{V}\} \in \mathcal{Y}$ where $\mathcal{Y} = \mathcal{Y}_1 \times ... \times \mathcal{Y}_N$ takes place when each $j \in \mathcal{V}$, first performs its local communication rule $\mu_j : \mathcal{Y}_j \to \mathcal{U}_j$ based on only y_j , and as soon as $u_{ne(j)}$ are collected, proceeds with the local estimation rule $\nu_j : \mathcal{Y}_j \times \mathcal{U}_{ne(j)} \to \mathcal{X}_j$.

Let $\gamma_j = (\mu_j, \nu_j)$ and $\gamma = (\gamma_1, ..., \gamma_N)$ denote the local rule of node j and the strategy of the network respectively. Let $\mathcal{M}_j$ and $\mathcal{N}_j$ denote the set of all possible communication and estimation rules respectively local to node j. Then, $\Gamma_j = \mathcal{M}_j \times \mathcal{N}_j$ for $\gamma_j \in \Gamma_j$ and the set of possible strategies given $\mathcal{G}$ is $\Gamma^{\mathcal{G}} = \Gamma_1 \times ... \times \Gamma_N$.

2.2. Problem Definition

As $(U, \hat{X}) = \gamma(Y)$, the joint process $(U, \hat{X}, X)$ has the joint density $p(u, \hat{x}, x; \gamma) = \int_{y\in\mathcal{Y}} \mathrm{d}y \; p(u, \hat{x}|x, y; \gamma)p(x, y)$ where "$;\gamma$" denotes that the distribution is specified by the processing strategy γ. Here $p(u, \hat{x}|x, y; \gamma) = \prod_{j=1}^N p(u_j, \hat{x}_j|y_j, u_{ne(j)}; \gamma_j)$ holds where $p(u_j, \hat{x}_j|y_j, u_{ne(j)}; \gamma_j) = p(u_j|y_j; \mu_j)p(\hat{x}_j|y_j, u_{ne(j)}; \nu_j)$ considering the causal online processing scheme correponding to $\mathcal{G}$ (described in Sec. 2.1). We note that the conditionals determined by local communication and estimation rules are $p(u_j|y_j; \mu_j) = \delta_{u_j, \mu_j(y_j)}$ and $p(\hat{x}_j|y_j, u_{ne(j)}; \nu_j) = \delta(\hat{x}_j - \nu_j(y_j, u_{ne(j)}))$ where $\delta_{i,j}$ and $\delta(x)$ are the Kronecker's and Dirac' s delta respectively.

Since the correspondance of $p(u, \hat{x}, x; \gamma)$ and γ are set, a cost function c which penalizes the estimation error of the pair $(x,\hat{x})$ and the communication load u, i.e. $c : \mathcal{U}\times\mathcal{X}\times\mathcal{X}\to\mathbb{R}$, yields an objective value for any strategy $\gamma \in \Gamma^{\mathcal{G}}$ given by the Bayesian risk $J(\gamma) = E\{c(u, x, \hat{x}); \gamma\}$ where the expectation is over $p(u, \hat{x}, x; \gamma)$. Given the constraints modelled with $\mathcal{G}$ and c, the best strategy for estimation is the solution to the optimization problem given by

$$\text{(P)} : \min J(\gamma) \;\; \text{, subject to } \; \gamma \in \Gamma^{\mathcal{G}} \tag{1}$$

2.3. Team Theoretic Iterative Solution

Team problems are involved in choosing best actions $\gamma_j \in \Gamma_j$ for $j = 1, ..., N$ with a single cost $J(\gamma_1, ..., \gamma_N)$. Concerned with minimization, when it is hard to find the global optimum, a useful relaxation is the Nash equilibrium $(\gamma_1^*, ..., \gamma_N^*)$ which satisfies

$$\gamma_j^* = arg \min_{\gamma_j\in\Gamma_j} J(\gamma_j, \gamma_{\backslash j}^*) \tag{2}$$

Algorithm 1 *Iterations converging to a pbp optimal strategy.*

0) (Initiate) $l = 0$, *choose* $\gamma^0 \in \Gamma$ *where* $\Gamma = \Gamma_1 \times ... \times \Gamma_N$;
1) (Update) $l = l + 1$;
For $j = 1, ..., N$
$\gamma_j^l = \arg\min_{\gamma_j\in\Gamma_j} J(\gamma_1^l, ..., \gamma_{j-1}^l, \gamma_j, \gamma_{j+1}^{l-1}, ..., \gamma_N^{l-1})$
2) (Check) If $J(\gamma^{l-1}) - J(\gamma^l) < \varepsilon$ *stop, else GO TO 1;*

for $j = 1, 2, ..., N$ where $\backslash j = \{1, 2, ..., N\}\backslash\{j\}$. $(\gamma_1^*, ..., \gamma_N^*)$ is also called a person by person (pbp) optimal solution [9]. It can easily be shown that Algorithm 1 converges to a pbp optimal strategy.

Problem (P) is NP-hard in the detection setting [7]. Considering a pbp optimal solution, provided that some reasonable assumptions hold, both the implied online processing and the update step of Algorithm 1 scales with the number of nodes. It is also possible to carry out this step in a message passing fashion. We follow this solution approach for estimation. These assumptions are

Assumption 1 (Conditional Independence): Noise processes are mutually independent yielding $p(x, y) = p(x) \prod_{i=1}^N p(y_i|x)$.

Assumption 2 (Measurement Locality): y_j is induced only by x_j for all $j \in \mathcal{V}$, i.e. $p(y_j|x) = p(y_j|x_j)$.

Assumption 3 (Separable Cost): The Bayesian cost function is of the form $c(u, \hat{x}, x) = c^d(\hat{x}, x) + \lambda c^c(u, x)$ where λ is a unit conversion coefficent which is the estimation error penalty equivalent to a unit communication cost.

Assumption 4 (Cost Locality): c^d and c^c are additive over nodes, i.e. $c(u, \hat{x}, x) = \sum_{j\in\mathcal{V}} c_j^d(\hat{x}_j, x_j) + \lambda \sum_{j\in\mathcal{V}} c_j^c(u_j, x_j)$.

Proposition (1): For Problem (P), if Assumptions 1-4 hold, $J(\gamma) = J_d(\gamma) + \lambda J_c(\gamma)$ and given a pbp optimal strategy $\gamma^* = (\gamma_1^*, ...\gamma_N^*)$ and fixing all local rules other than the j^{th}, the j^{th} optimal rule given by Eq.(2) reduces to local communication and estimation rules $\mu_j^*(Y_j)$ and $\nu_j^*(Y_j, U_{ne(j)})$ given by

$$\arg\min_{u_j\in\mathcal{U}_j} \int_{\mathcal{X}_j} \mathrm{d}x_j p(x_j)p(Y_j|x_j)[\lambda c_j^c(u_j, x_j) + C_j^*(u_j, x_j)] \tag{3}$$

$$\arg\min_{\hat{x}_j\in\mathcal{X}_j} \int_{\mathcal{X}_j} \mathrm{d}x_j p(x_j)p(Y_j|x_j)P_j^*(U_{ne(j)}|x_j)c_j^d(\hat{x}_j, x_j) \tag{4}$$

respectively where $\forall u_{ne(j)} \in \mathcal{U}_{ne(j)}$

$$P_j^*(u_{ne(j)}|x_j) = \int_{\mathcal{X}_{ne(j)}} \mathrm{d}x_{ne(j)} p(x_{ne(j)}|x_j) \prod_{i\in ne(j)} P_{i\to j}^*(u_{i\to j}|x_i) \tag{5}$$

with terms regarding influence of $i \in ne(j)$ on j given by $P_{i\to j}^*(u_{i\to j}|x_i) = \sum_{u_i\backslash u_{i\to j}} p(u_i|x_i; \mu_i^*)$, $\forall u_{i\to j} \in \mathcal{U}_{i\to j}$ where $p(u_i|x_i; \mu_i^*) = \int_{\mathcal{Y}_i} \mathrm{d}y_i p(y_i|x_i)p(u_i|y_i; \mu_i^*)$. In addition $\forall u_j \in \mathcal{U}_j$

$$C_j^*(u_j, x_j) = \sum\nolimits_{i\in ne(j)} C_{i\to j}^*(u_{j\to i}, x_j) \tag{6}$$

holds with terms regarding the influence of j on $i \in ne(j)$ given by

$$C_{i\to j}^*(u_{j\to i}, x_j) = \int_{\mathcal{X}_{ne(i)\backslash j}} dx_{ne(i)\backslash j} \int_{\mathcal{X}_i} dx_i \, p(x_{ne(i)\backslash j}, x_i|x_j) \times \sum_{u_{ne(i)\backslash j}} \prod_{j'\in ne(i)\backslash j} P_{j'\to i}^*(u_{j'\to i}|x_{j'}) I_i^*(u_{ne(i)}, x_i; \gamma_i^*) \tag{7}$$

such that

$$I_i^*(u_{ne(i)}, x_i; \nu_i^*) = \int_{\mathcal{Y}_i} \mathrm{d}y_i \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i c_i^d(\hat{x}_i, x_i)p(\hat{x}_i|y_i, u_{ne(i)}; \nu_i^*) \times p(y_i|x_i) \tag{8}$$

Proof: Due to lack of space we skip the proof here but an analogous version of this proposition has been proved for the detection problem [8]. The above expressions can be obtained from this version by replacing summations over $\mathcal{X}_j$s with integrations, changing the order of operators appropriately and assuming that the links are error-free.

With the proposition above, given a pbp optimal strategy, we obtain communication and estimation rules local to node j in terms of the remaining in a variational form. Considering $P_{i\to j}^*(u_{i\to j}|x_i)$

Algorithm 2: *Iterations converging to a person by person optimal decentralized estimation strategy for Problem (P).*

0) (Initiate) $l = 0$, *choose* $\gamma^0 \in \Gamma^{\mathcal{G}}$;
1) (Update) $l = l + 1$;
For $i = 1, ..., N$, *Compute* $\{P^l_{i\to j}(u_{i\to j}|x_j)\}_{j\in ne(i)}$;
For $i = 1, ..., N$, *Update* ν^l_i, *compute* $\{C^l_{i\to j}(u_{j\to i}, x_j)\}_{j\in ne(i)}$;
For $i = 1, ..., N$, *Update* μ^l_i;
2) (Check) If $J(\gamma^{l-1}) - J(\gamma^l) < \epsilon$ *stop, else GO TO (1);*

for $i \in ne(j)$, $P^*_j(u_{ne(j)}|x_j)$ is the likelihood of x_j given $u_{ne(i)}$. Eq.s(6)-(8) reveal that $C^*_j(u_j, x_j)$ is the total expected cost induced on the neighbors by u_j, i.e. $E\{c(u_{ne(j)},\hat{x}_{ne(j)},x_{ne(j)})|u_j,x_j\}$. Hence, we conclude that the j^{th} optimal communication rule selects the message that results with a minimum contribution to the overall cost and also noting that $p(x_j)p(y_j|x_j)P(u_{ne(j)}|x_j) \propto p(x_j|y_j, u_{ne(j)})$ holds under Assumptions 1-4, the optimal estimation rule selects $\hat{x}_j$ that yields minimum expected penalty given y_j and $u_{ne(j)}$.

The right hand sides of Eq.s(5)-(8) can be treated as operators valid for any set of local rules. Hence it is possible to specify the update step of Algorithm 1 for Problem (P) and obtain Algorithm 2. The objective value at l^{th} step is easily found to be

$$J(\gamma^l) = \sum_{i\in\mathcal{V}} G^d_i(\nu^l_i) + \lambda \sum_{i\in\mathcal{V}} G^c_i(\mu^l_i) \tag{9}$$

where $G^d_i(\nu^l_i) = \sum_{u_{ne(i)}}\int_{\mathcal{X}_i} dx_i p(x_i) P^{l+1}_i(u_{ne(i)}|x_i) I_i(u_{ne(i)}, x_i; \nu^l_i)$ and $G^c_i(\mu^l_i) = \sum_{u_i}\int_{\mathcal{X}_i} dx_i c^c_i(u_i, x_i)p(x_i)p(u_i|x_i;\mu^l_i)$ in terms of the expressions discussed above.

It is possible to carry out the update step of Algorithm 2 in a message passing fashion where in the first pass each node i sends $P^l_{i\to j}$ to $j \in ne(i)$ and upon reception of these terms from all neighbors, updates $P^l_i(u_{ne(i)}|x_i)$ and ν^l_i accordingly. In the second pass node i sends $C^l_{i\to j}$ to $j \in ne(i)$ and as soon as it receives all the cost messages from neighbors, μ^l_i is updated.

3. MONTE CARLO APPROXIMATED ITERATIONS

For problem (P), Algorithm 2 yields a pbp optimal solution in principle. The operators required in the update step and implied by Eq.s(5)-(8) as well as the pbp optimal local rules given by Eq.s(3)-(4) do not have closed form solutions in general for which we propose particle representations and corresponding approximate computational schemes through MC integration methods presented in Section 3.1. In Section 3.2 we progressively apply them and obtain an approximation to the local rule described in Proposition (1).

3.1. Monte Carlo Integration

Consider $i = \int_{\mathcal{X}} dx\, p(x)f(x)$, where $p(x)$ is a probability density for X such that a realization x satistfies $x \in \mathcal{X}$. In the conventional MC method, given M independent samples, i.e. $x^{(m)} \sim p(x)$ for $m = 1, ..., M$, i is estimated with $\hat{i}_M = \frac{1}{M}\sum_{k=1}^M f(x^{(k)})$ which exhibits almost sure convergence. If we are able to maintain $x^{(m)} \sim g(x)$ for $m = 1, ..., M$ instead, the Importance Sampling (IS) method proposes $\hat{i}_M = \frac{1}{M}\sum_{k=1}^M \omega_{(k)} f(x^{(k)})$ where $\omega_{(k)} = p(x^{(k)})/g(x^{(k)})$ which also converges to i almost surely if the support of g is covered by that of f. When a small number of weights dominate, $\hat{i}_M = \left(1/\sum_{k=1}^M \omega_{(k)}\right)\sum_{k=1}^M \omega_{(k)} f(x^{(k)})$ is preferable although it is slightly biased for small M [10].

3.2. Iterative MC Optimization Scheme

Considering Proposition (1), we proceed in three steps;

Step 1 We replace the integrals appearing in the local rule expressions given in Eq.s (3) and (4) with conventional MC approximations, i.e. given $x^{(m)}_j \sim p(x_j)$ for $m = 1, ..., M$, with $(1/M)\sum_{m=1}^M p(Y_j|x^{(m)}_j)[\lambda c^c_j(u_j, x^{(m)}_j) + C^*_j(u_j, x^{(m)}_j)]$ and $(1/M)\sum_{m=1}^M p(Y_j|x^{(m)}_j) P^*_j(U_{ne(j)}|x^{(m)}_j) c^d_j(\hat{x}_j, x^{(m)}_j)$ respectively.

Step 2 Both P^*_j and C^*_j are required to be known $\forall u_{ne(j)} \in \mathcal{U}_{ne(j)}$ and $\forall u_j \in \mathcal{U}_j$ respectively for $\{x^{(m)}_j\}_{m=1}^M$ in Step 1. Assuming that $\{C^*_{i\to j}(u_{j\to i}, x^{(m)}_j)\}_{m=1}^M$ are known $\forall i \in ne(j)$ and $\forall u_{j\to i} \in \mathcal{U}_{j\to i}$ Eq.(6) directly applies. Given $\{P^*_{i\to j}(u_{i\to j}, x^{(m)}_i)\}_{m=1}^M \forall u_{i\to j} \in \mathcal{U}_{i\to j}$ where $x^{(m)}_i \sim p(x_i), m = 1, ..., M$ and noting that $\{x^{(m)}_i\}_{i\in ne(j)} \sim \prod_{i\in ne(j)} p(x_i)$ an IS approximation to $P^*_j(u_{ne(j)}|x^{(m)}_j)$ given by Eq.(5) is through weights $\omega^{(m)(m')}_j = p(x^{(m')}_{ne(j)}|x^{(m)}_j)/\prod_{i\in ne(j)} p(x^{(m')}_i)$

$$\tilde{P}^*_j(u_{ne(j)}|x^{(m)}_j) = \frac{1}{\sum_{m'=1}^M \omega^{(m)(m')}_j} \sum_{m'=1}^M \omega^{(m)(m')}_j \prod_{i\in ne(j)} P^*_{i\to j}(u_{i\to j}|x^{(m')}_i)$$

Step 3 In this step we approximate the node to node terms. For $i \in ne(j)$, $P^*_{i\to j}(u_{i\to j}, x^{(m)}_i)$ is a marginalization of $p(u_i|x^{(m)}_i; \mu^*_i)$. For $m = 1, ..., M$ an IS approximation to this conditional distribution is possible through $y^{(p)}_i \sim p(y_i)$, $p = 1, ..., P$ with weights $\omega^{(m)(p)}_i = p(y^{(p)}_i|x^{(m)}_i)/p(y^{(p)}_i)$ as

$$\tilde{p}(u_i|x^{(m)}_i; \mu^*_i) = \frac{1}{\sum_{p=1}^P \omega^{(m)(p)}_i} \sum_{p=1}^P \omega^{(m)(p)}_i \delta_{u_i, \mu^*_i(y^{(p)}_i)} \tag{10}$$

Considering the conditionals in Section 2.2, an IS approximation to $I^*_i(u_{ne(i)}, x^{(m)}_i; \nu^*_i), \forall u_{ne(i)} \in \mathcal{U}_{ne(i)}$ and for $m = 1, ..., M$ using the already generated sample set $\{y^{(p)}_i\}_{p=1}^P$ and the IS weights above is

$$\tilde{I}^*_i(u_{ne(i)}, x^{(m)}_i; \nu^*_i) = \frac{1}{\sum_{p=1}^P \omega^{(m)(p)}_i} \sum_{p=1}^P \omega^{(m)(p)}_i c^d_i(\nu^*_i(y^{(p)}_i, u_{ne(i)}), x^{(m)}_i)$$

Next we consider Eq.(7) for which assuming that $\forall j' \in ne(i)\backslash j$, $\{P^*_{j'\to i}(u_{j'\to i}, x_{j'}{}^{(m)})\}_{m=1}^M$ $\forall u_{j'\to i} \in \mathcal{U}_{j'\to i}$ are given where $x^{(m)}_{j'} \sim p(x_{j'})$ and noting that $x^{(m)}_{ne(i)\backslash j} \sim \prod_{j'\in ne(i)\backslash j} p(x_{j'})$ where $x^{(m)}_{ne(i)\backslash j} \triangleq \{x^{(m)}_{j'}\}_{j'\in ne(i)\backslash j}$ an IS approximation $\forall u_{j\to i} \in \mathcal{U}_{j\to i}$ and for $m = 1, ..., M$ is

$$\tilde{C}^*_{i\to j}(u_{j\to i}, x^{(m)}_j) = \sum_{u_{ne(i)\backslash j}} \frac{1}{\sum_{m'=1}^M \omega^{(m)(m')}_i} \sum_{m'=1}^M \omega^{(m)(m')}_i \times \prod_{j'\in ne(i)\backslash j} P^*_{j'\to i}(u_{j'\to i}|x^{(m')}_{j'}) \tilde{I}^*_i(u_{ne(i)}, x^{(m')}_i; \nu^*_i)$$

where $\omega^{(m)(m')}_i = p(x^{(m')}_{ne(i)\backslash j}, x^{(m')}_i|x^{(m)}_j)/p(x^{(m')}_i) \prod_{j'\in ne(i)\backslash j} p(x^{(m')}_{j'})$.

The above steps render an approximated counterpart of Proposition (1) resulting $\tilde{\gamma}^*_j \approx \gamma^*_j$. When applied for all nodes $i \in \mathcal{V}$, they provide computationally feasible approximations for the update step of Algorithm (2), which in turn implies a MC optimization scheme yielding $\tilde{\gamma}^*$ given by Algorithm (3). For checking convergence, an approximation $\tilde{J}(\tilde{\gamma}^l) \approx J(\gamma^l)$ is immediate through substituting $\tilde{G}^d_i(\tilde{\nu}^l_i) = \sum_{u_{ne(i)},m} \tilde{P}^{l+1}_i(u_{ne(i)}|x^{(m)}_i)\tilde{I}^l_i(u_{ne(i)}, x^{(m)}_i; \tilde{\nu}^l_i)$ and $\tilde{G}^c_i(\tilde{\mu}^l_i) = \sum_{u_i,m} c^c_i(u_i, x^{(m)}_i)p(u_i|x^{(m)}_i; \tilde{\mu}^l_i)$ in Eq.(9). Hence, after selecting an initial strategy and generating $\{\{x^{(m)}_j\}_{m=1}^M\}_{j=1}^N$ where $x^{(m)}_j \sim p(x_j)$ and $\{\{y^{(p)}_j\}_{p=1}^P\}_{j=1}^N$ where $y^{(p)}_j \sim p(y_j)$, Algorithm (3) approaches an approximately pbp optimal strategy constrained by the undirected graph $\mathcal{G}$.

4. EXAMPLE

Consider a DE network represented with the UG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in Figure (1a) with $\mathcal{U}_{i\to j} = \{0, 1, 2\}$ $\forall(i, j) \in \mathcal{E}$. For each node i, $c^c_i(u_i, x_i) = \sum_{j\in ne(i)} c(u_{i\to j})$ where $c(u_{i\to j}) = 0$ if $u_{i\to j} = 0$

Algorithm 3: *Iterative MC algorithm that converges to an approximate pbp optimal decentralized strategy.*

0) (Initiate) $l = 0$, *choose* $\gamma^0 \in \Gamma^{\mathcal{G}}$;
1) (Update) $l = l + 1$;
For $i = 1, ..., N$, *Compute* $\{\{\tilde{P}^l_{i\to j}(u_{i\to j}|x_j^{(m)})\}_{m=1}^M\}_{j\in ne(i)}$;
For $i = 1, ..., N$
Update $\tilde{\nu}_i^l$, *compute* $\{\{\tilde{C}^l_{i\to j}(u_{j\to i}, x_j^{(m)})\}_{m=1}^M\}_{j\in ne(i)}$;
For $i = 1, ..., N$, *Update* $\tilde{\mu}_i^l$;
2) (Check)If $|\tilde{J}(\tilde{\gamma}^{l-2}) - \tilde{J}(\tilde{\gamma}^{l-1})| - |J(\tilde{\gamma}^{l-1}) - \tilde{J}(\tilde{\gamma}^l)| > \varepsilon$ *GO TO (1)*;
else $\tilde{\gamma}^* = \tilde{\gamma}^l$, *STOP*;

and $c(u_{i\to j}) = 1$ otherwise. Hence J_c is the total expected link use rate (LUR) in bits. The estimation error penalty is $c_i^d = (x_i - \hat{x}_i)^2$ and J_d is the total mean squared error (MSE).

Subject to estimation is a multivariate Gaussian random field, i.e. $x \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_x)$, which is Markov with respect to the graph in Figure(1b). We choose $\mathbf{C}_x$ accordingly as

$$\mathbf{C}_x = \begin{bmatrix} 2 & 1.125 & 1.5 & 1.125 \\ 1.125 & 2 & 1.5 & 1.125 \\ 1.5 & 1.5 & 2 & 1.5 \\ 1.125 & 1.125 & 1.5 & 2 \end{bmatrix} \quad (11)$$

The j^{th} field of x is associated with platform j and the noise processes $\{n_j\}_{j\in\mathcal{V}}$ are additive, mutually independent and Gaussian, i.e. $n_j \sim \mathcal{N}(0, \sigma_n^2)$ where $\sigma_n^2 = 0.5$, yielding an SNR of 6dB for each sensor. For each platform j, the initial local estimation rule is the myopic mimimum MSE estimator which is based only on y_j, i.e. $\nu_j^0(y_j, u_{ne(j)}) = \int_{-\infty}^{\infty} \mathrm{d}x_j\, x_j p(x_j|y_j)$, and the communication rule is a threshold rule quantizing y_j, i.e. $\mu_i^0(y_i, u_{ne(i)}) = 1, 0$ and 2 for $y_i < -2\sigma_n$, $-2\sigma_n \leq y_i < 2\sigma_n$ and $y_i \geq 2\sigma_n$ respectively.

The performance point (J_c, J_d) of the converged strategy vary with λ. For $\lambda \geq \lambda^*$, no transmission with myopic estimation rules achieve the minimum cost which is also a pbp optimal. Hence, λ^* admits an interpretation of being the maximum price per bit that the system affords to decrease the estimation penalty. We approximate the performance curve of solutions as we increase λ from 0 which is an approximate quantification for the tradeoff between the cost of estimation errors and communication.

In Figure (1c) we present these pairs, i.e. $(\tilde{J}_c, \tilde{J}_d)$, for different choices of λ and $|\mathcal{U}_{i\to j}|$s. The upper and lower limits are MSEs corresponding to the myopic rule and the centralized optimal rule[1] respectively. $(\tilde{J}_c, \tilde{J}_d)$ points for the 1-bit selective communication scheme reveal that altough the transmission has no cost for $\lambda = 0$, the total link use rate is only slightly higher than 50% of the total 6 bits indicating that the information from receiving no messages is successfully utilized. Moreover, the MSE performance is closer to that of the centralized scheme than the myopic scheme. The communication stops for $\lambda^* \approx 0.3$. Approximate performance points for 2-bits case present the decrease in MSE for the same network load as we increase the link capacities for small values of λ which is competetive with that of the centralized rule.

5. CONCLUSION

We have considered the design of a decentralized estimation network constrained with an undirected communication graph in a Bayesian framework that captures costs due to both estimation errors and transmissions. Adopting a recent scheme for detection networks which proposes a solution utilizing team decision theory we have extended the set of constraints considered by the conventional approaches for the decentralized estimation problem. In principle, the solution is optimal in a person by person sense and achieved iteratively. We have proposed particle representations and approximate computational schemes utilizing Monte Carlo methods for the operators we encounter in the iterative algorithm, which are impossible to evaluate exactly in practice in general. We maintain scalability with the number of nodes as well as the size of the sample sets. This efficiency enables us to approximately quantify the tradeoff between estimation accuracy and communication cost through the performance curves.

[1] For $c(x, \hat{x}) = (x - \hat{x})^T(x - \hat{x})$, the optimal centralized estimate is the mean of $p(x_1, ..., x_4|y_1, ..., y_4)$ which yields a minimum of $J_c = 3Q$ bits where Q is the number of bits used to quantize y_j before transmission.

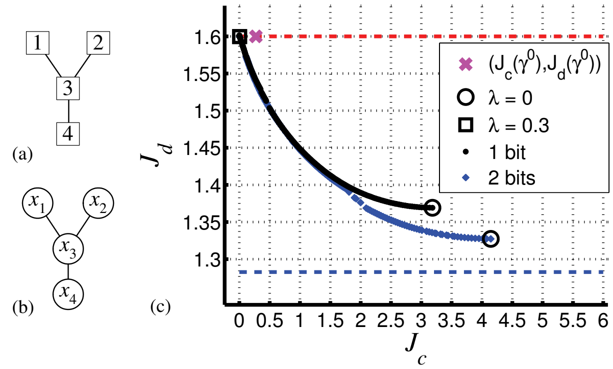

Fig. 1. (a) UG topology of the DE network, (b) Markov Random Field representation of X, (c) Approximate points of the performance curves while λ is increased from 0 with 0.001 steps, for the example scenario.

6. REFERENCES

[1] J. A. Gubner, "Distributed estimation and quantization," *IEEE Trans. on Info. Theory*, vol. 6, pp. 1456–1459, July 1993.

[2] W.-M. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," *IEEE Trans. on Communications*, vol. 41, pp. 1602–1605, Nov. 1993.

[3] A. Ribeiro, I. D. Schizas, J.-J. Xiao, G. B. Giannakis, and Z. Q. Luo, *Wireless Sensor Networks: Signal Processing and Communication Perspectives*, chapter Distributed Estimation under Bandwidth and Energy Constraints, J. W. & Sons, Ltd., 2007.

[4] M. Cetin, L. Chen, J. W. Fisher III, A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, "Distributed fusion in sensor networks," *IEEE Signal Processing Magazine*, vol. 23, pp. 42–55, July 2006.

[5] A. T. Ihler, J. W. Fisher III, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *Jour. of Mach. Learning Research*, vol. 6, pp. 905–936, May 2005.

[6] Murat Üney and Müjdat Çetin, "An efficient Monte Carlo approach for optimizing communication constrained decentralized estimation networks," in *The* 17^{th} *EUSIPCO*. EURASIP, Aug. 2009, To appear.

[7] O. P. Kreidl and A. S. Willsky, "Decentralized detection in undirected network topologies," in *Proc. of SSP Workshop*. IEEE, Aug. 2007.

[8] O. P. Kreidl, *Graphical Models and Message-Passing Algorithms for Network- Constrained Decision Problems*, Ph.D. thesis, MIT Dept. of EECS, 2008.

[9] Y.-C. Ho and K.-C. Chu, "Team decision theory and information structures in optimal control problems," *IEEE Trans. on Auto. Control*, vol. 17, pp. 15–22, Feb. 1972.

[10] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2^{nd} edition, 2004.