Optimization of Decentralized Random Field Estimation Networks Under Communication Constraints through Monte Carlo Methods

Murat Üney[1,*], Müjdat Çetin

Faculty of Engineering and Natural Sciences, Sabancı University, Orhanlı-Tuzla 34956 İstanbul, Turkey

Abstract

We propose a new methodology for designing decentralized random field estimation schemes that takes the tradeoff between the estimation accuracy and the cost of communications into account. We consider a sensor network in which nodes perform bandwidth limited two-way communications with other nodes located in a certain range. The in-network processing starts with each node measuring its local variable and sending messages to its immediate neighbors followed by evaluating its local estimation rule based on the received messages and measurements. Local rule design for this two-stage strategy can be cast as a constrained optimization problem with a Bayesian risk capturing the cost of transmissions and penalty for the estimation errors. A similar problem has been previously studied for decentralized detection. We adopt that framework for estimation, however, the corresponding optimization schemes involve integral operators that are impossible to evaluate exactly, in general. We employ an approximation framework using Monte Carlo methods and obtain an optimization procedure based on particle representations and approximate computations. The procedure operates in a message-passing fashion and generates results for any distributions if samples can be produced from, e.g., the marginals. We demonstrate graceful degradation of the estimation accuracy as communication becomes more costly.

Keywords: Decentralized estimation, communication constrained inference, random fields, message passing algorithms, Monte Carlo methods, wireless sensor networks

2000 MSC: 94A99

☆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 by a Turkish Academy of Sciences Distinguished Young Scientist Award.

*Corresponding author.

Email addresses: muratuney@sabanciuniv.edu (Murat Üney), mcetin@sabanciuniv.edu (Müjdat Çetin)

[1]This research was done at Sabancı University, İstanbul, Turkey. Murat Üney is currently a Research Fellow at the Institute for Digital Communications (IDCOM), the University of Edinburgh, Edinburgh EH9 3JL, UK (e-mail:M.uney@ed.ac.uk, tel:+44 131 650 5659, fax:+44/0 131 650 6554).

1. Introduction

Wireless sensor networks have been a promising technology for deploying a large number of sensor platforms over a region to gather dense spatial samples of a physical phenomenon [1]. Applications including environmental monitoring, structural monitoring [2] and precision agriculture [3] benefit from wirelessly networking these platforms in an ad-hoc fashion which can also collect measurements in possibly multiple modes induced by multiple quantities of interest. There are challenges in design because the sensor platforms have limited computational and energy resources and the links over which they can communicate are bandwidth (BW) limited. The dispersed nature of the system necessitates some communications for processing the measurements, however, the energy cost of transmitting bits is usually greater than that for computing them [4]. Therefore, it is crucial for the feasibility of a sensor network to take the estimation-communication tradeoffs into account while performing collaborative "online" (or, in–network) processing of the measurements in the network [5].

In this context, we are concerned with designing decentralized processing schemes for random field estimation under a set of communication constraints. In the network structure we consider, the platforms perform local communication with their neighbors located within a certain range and form a connected ad-hoc network with BW limited links. We are particularly interested in the tradeoff between the estimation accuracy and the cost of trasmissions given the link topology. Transmission costs might include the energy cost of communications through, e.g., an energy dissipation model for transmitting and receiving k bits at a distance of d meters [6].

Subject to estimation is a set of spatial random variables that exhibit a correlation structure. Examples of physical phenomena that can be modeled with such random fields include turbulent flow (Chp. 12 of [7]) and geostatistical data [8] such as temperature measurements over a field (Chp. 1 of [9]). There is a variety of lines of investigation on random field estimation with sensor networks. In-network processing schemes based on adaptive hierarchies (e.g., [10]), a designated fusion center (FC) receiving quantized measurements (e.g., [11]), and iterations involving FC feedback [12] have been considered. These treatments cannot pose an in-network strategy design problem that explicitly takes the tradeoffs into account and are not decentralized in that not all of the nodes contribute to the estimation task but only one or more FCs. Estimation of dynamic random fields through Kalman-Bucy filtering (KBF) is considered in [13] and [14]. In particular, [14] introduces a distributed realization of the KBF, whereas [13] considers an FC that collects measurements from sensors after finding a reduced model whereby a subset of the sensors are queried based on a surrogate communication costs and an estimation penalty. Our problem setting differs in that we are concerned with completely decentralized strategies and, on a static problem, consider the trade-off between the estimation accuracy and the communication load of the network.

Decentralized estimation in sensor networks has also been studied using probabilistic graphical models (see, e.g., [15] and the references therein). In this approach, a probabilistic dependency graph of the random field is mapped onto the communication topology. The in-network processing strategy then becomes a message passing algorithm which communicates probability distributions. However, model approximations together with message coding and censoring to facilitate low-energy digital transmissions complicate the performance analysis [16]. As a result, it is not straightforward to state a design problem that takes the network topology and the communication cost into account using this perspective [17].

We consider a class of in-network processing strategies which operate over an undirected communication topology and yield a rigorous communication constrained design problem through a tractable Bayesian risk. In particular, the platforms specify the vertex set and the undirected edges represent bi-directional communication links with finite alphabets sizes of which are related to the BWs. The nodes estimate a (set of) random variable(s) possibly related to a random field model based on the platform locations through a two-stage procedure: In the first stage, each node makes a measurement and produces messages to its neighbors using its communication rule. In the second stage nodes estimate their associated random variable(s), based on both the incoming messages and their measurements. The design problem involves finding the communication and estimation rules for the nodes and it is in the form of a constrained optimization problem in which the objective function is a Bayesian risk that penalizes both estimation errors and the transmissions, and the feasible set of strategies is constrained by the corresponding graph representation that captures the availability and the capacity of the links.

A similar problem has been recently studied in the context of decentralized *detection* [18] based upon the results for another class of strategies – those over directed acyclic graphs (DAGs)(see also [19]). One appealing feature of this approach is that the solution to the design problem can be realized as a message passing algorithm which fits well into the distributed system requirements of a sensor network. We have considered the design of decentralized *estimation* strategies over DAGs in [20], and introduced an approximation framework through Monte Carlo (MC) methods in order to overcome the difficulties arising from the fact that the variables of concern take values from nondenumerable sets in the estimation case. This paper differs from recent work taking a similar distributed inference perspective in that we consider estimation problems (rather than detection problems as in [18, 19, 21]) over undirected graphs (UGs) (rather than DAGs as in [20]).

The contribution of this paper is an adoption of the aforementioned approximation framework for the class of (decentralized) two-stage estimation strategies *over UGs* which we believe is a good match for random field estimation scenarios. Doing that, we transform a Team Decision Theoretic (TDT) iterative strategy optimization to a computationally feasible MC optimization algorithm which employs nonparametric representations of the underlying distributions. We also maintain the benefits of the TDT solution and, as a result, our approach features the following: First, this framework enables us to consider a broad range of communication and computation structures for the design of decentralized estimation networks. Second, in the case that a dual objective is selected as a weighted-sum of the estimation performance and the cost of communications, a graceful degradation of the estimation accuracy is achieved as communication becomes more costly. The resulting pareto-optimal curve enables a quantification of the tradeoff of concern. Under reasonable assumptions, the optimization procedure scales with the number of platforms as well as the number of variables involved. Moreover, it can be realized as a message passing algorithm which is an appropriate computational structure for network self-organization. The MC optimization scheme we propose features scalability with the cardinality of the sample sets required and can produce results for any set of distributions provided that independent samples can be generated from, e.g., the marginals.

In Section 2, we introduce the design problem in a constrained optimization setting, and then we describe the Team Decision Theoretic investigation of its solution in Section 3. We present our MC optimization framework for two-stage in-network processing strategies over UGs in Section 4. Then, we demonstrate the aforementioned features through several examples in Section 5[2]. Finally, we provide concluding remarks in Section 6.

2. Problem Definition

In this section, we start introducing the problem setting with some basic definitions. Then, in Section 2.1 we present the two-stage in-network processing scheme over an undirected communication topology. In Section 2.2, we state the strategy design problem as a constrained optimization problem taking into account the communication constraints. This problem is to be solved *offline*, i.e., before processing the observations.

[2] The preliminary results of the proposed scheme appear in [22].

We consider N sensor platforms dispersed over a region. Each node can establish communication links with some of the other nodes within its communication range. These links are bi-directional and the communications structure can be represented by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in which each platform is associated with a node $v \in \mathcal{V}$. An edge $(i, j) \in \mathcal{E}$ corresponds to a finite capacity one-way link from platform

Table 1: Nomenclature for the in-network processing strategy.

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	Undirected graph of the set of nodes $\mathcal{V}$ and the set of bi-directional communication links $\mathcal{E}$.
X_j	Random variable associated with node j.
Y_j	Random variable modeling the measurement taken by node j.
(X, Y)	Joint random variable modeling the estimation problem.
x_j	Realization of X_j in the joint event.
y_j	Measurement taken by node j.
$\hat{x}_j$	Estimated value of x_j drawn by node j.
$u_{i\to j}$	Message symbol from node i to j.
$\mathcal{U}_{i\to j}$	Set of admissible symbols from node i to j.
$\overrightarrow{u}_j$	Vector of messages from node j to its neighbors.
$\overleftarrow{u}_j$	Vector of messages to node j from its neighbors.
$\mu_j(y_j)$	Communication rule of node j outputting $\overrightarrow{u}_j$.
$\mathcal{M}_j^{\mathcal{G}}$	Space of feasible communication rules for node j.
ν_j	Estimation rule of node j outputting $\hat{x}_j$ given $(y_j, \overleftarrow{u}_j)$.
$\mathcal{N}_j^{\mathcal{G}}$	Space of feasible estimation rules for node j.
γ_j	The local rule pair (μ_j, ν_j) node j.
$\Gamma_j^{\mathcal{G}}$	Space of feasible local rule pairs for node j in $\mathcal{G}$.
γ	In–network processing strategy as a concatenation of all local rules.
$\Gamma^{\mathcal{G}}$	Space of all feasible strategies over $\mathcal{G}$.
$c(u, x, \hat{x})$	Cost of the communication vector u and the pair $(x, \hat{x})$.
$J(\gamma)$	Bayesian risk of γ.

i to j. The bi-directionality is captured by using a UG representation in which $(i, j) \in \mathcal{E} \iff (j, i) \in \mathcal{E}$. A particular example of such a network can be seen in Figure 5(a) in Section 5.3.

On the edge (i, j), node i transmits a symbol $u_{i \to j}$ from the set of admissible symbols $\mathcal{U}_{i \to j}$. For example, in order to model a link with capacity $\log_2 d_{ij}$ bits, one can select $\mathcal{U}_{i \to j}$ such that $\left|\mathcal{U}_{i \to j}\right| = d_{ij}$. In order to represent the "no transmission" event in censoring or selective communication schemes, one can insert an additional symbol into $\mathcal{U}_{i \to j}$ such as 0. We note that, as both (i, j) and $(j, i) \in \mathcal{E}$, the variables $u_{j \to i}$ and $u_{i \to j}$ are symbols in opposite directions over the same link.

Associated with each sensor platform is a set of variables modeling, e.g., the temperature, humidity, or the flow vector at possibly the position of the platform. Let us denote a concatenation of variables associated with node j by X_j and the set it takes values from by $\mathcal{X}_j$. In principle, there is no restriction on the dimensionality of $\mathcal{X}_j$, i.e., $\dim(\mathcal{X}_j) \geq 1$. All random variables to be estimated can be represented with a concatenation $X = (X_1, X_2, ..., X_N)$ which takes values from $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times ... \times \mathcal{X}_N$. For example, for real valued random variables, $\mathcal{X}_j = \mathbb{R}$ and $\mathcal{X} = \mathbb{R}^N$. It is worth reminding that, in the detection setting, $\mathcal{X}_j$s are $M < \infty$ element sets for M-ary detection.

Node j collects measurements Y_j using its onboard sensors. $Y_j \in \mathcal{Y}_j$ where $\mathcal{Y}_j$ is nondenumerable, as well. All observations collected by the network is denoted by $Y = (Y_1, Y_2, ..., Y_N)$ and resides in $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times ... \times \mathcal{Y}_N$.

The probabilistic model underlying the estimation problem is represented by the random variable pair (X, Y). It is characterized by the joint cumulative distribution function $P_{X,Y}(x, y)$ with the density $p_{X,Y}(x, y)$ for a realization $(x, y) = (x_1, ..., x_N, y_1, ..., y_N)$.

2.1. Two-stage in-network processing strategy over undirected graphs

Suppose we are given a UG communication topology $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The neighbors of node j is given by $ne(j) \triangleq \{i \mid (i, j) \in \mathcal{E} \wedge (j, i) \in \mathcal{E}\}$. Let us denote the set of outgoing messages from node j to its neighbors by $\overrightarrow{u}_j \triangleq \{u_{j \to i} \mid i \in ne(j)\}$. Then, $\overrightarrow{u}_j$ takes values from $\overrightarrow{\mathcal{U}}_j = \bigotimes_{i \in ne(j)} \mathcal{U}_{j \to i}$ where $\otimes$ denotes consecutive Cartesian products[3]. Being at the receiving end of the links from its neighbors, node j collects the incoming messages denoted by $\overleftarrow{u}_j \triangleq \{u_{i \to j} \mid i \in ne(j)\}$ and take values from $\overleftarrow{\mathcal{U}}_j = \bigotimes_{i \in ne(j)} \mathcal{U}_{i \to j}$. The messages across the network are similarly given by $u \triangleq \{u_{i \to j} \mid (i, j) \in \mathcal{E}\}$ and reside in $\mathcal{U} \triangleq \bigotimes_{(i,j) \in \mathcal{E}} \mathcal{U}_{i \to j}$.

At this point, it is worthwhile to point out that we implicitly assume the links in $\mathcal{G}$ are error free so that the symbols transmitted (or lack thereof) from neighbors are exactly restored at the receiving end. This

[3] In other words, e.g., $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$ and $\mathcal{X} = \bigotimes_{i \in \{1,2,3\}} \mathcal{X}_i$ are synonymous.

is for the sake of simplicity throughout the article and it is indeed possible to accommodate an unreliable channel model capturing link errors and packet losses possibly due to noise and interference in this network model [18][4].

We continue our discussion by specifying a two-stage operation that ensures a causal online processing without deadlocks: In the first stage, having observed $y_j \in \mathcal{Y}_j$, node j evaluates its local communication rule defined by $\mu_j : \mathcal{Y}_j \to \overrightarrow{\mathcal{U}}_j$ and produces outgoing messages to its neighbors[5]. After receiving all the messages from its neighbors, node j performs the second stage in which it evaluates its estimation rule given by $\nu_j : \mathcal{Y}_j \times \overleftarrow{\mathcal{U}}_j \to \mathcal{X}_j$ to draw an inference on the value X_j takes based on the observation y_j and the incoming messages $\overleftarrow{u}_j$ from neighboring nodes. Hence, the local rule of node j is a pair given by $\gamma_j = (\mu_j, \nu_j)$. The objective of designing γ_j is the topic of Section 2.2.

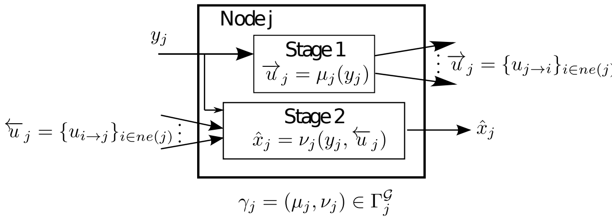
Based on the previous definitions, the space of all first-stage (communication) rules is defined as $\mathcal{M}_j^{\mathcal{G}} \triangleq \{\mu_j \mid \mu_j : \mathcal{Y}_j \to \overrightarrow{\mathcal{U}}_j\}$ and the second-stage (estimation) rule space is given by $\mathcal{N}_j^{\mathcal{G}} \triangleq \{\nu_j \mid \nu_j : \mathcal{Y}_j \times \overleftarrow{\mathcal{U}}_j \to \mathcal{X}_j\}$. Consequently, the space of rules local to node j is given by $\Gamma_j^{\mathcal{G}} \triangleq \mathcal{M}_j^{\mathcal{G}} \times \mathcal{N}_j^{\mathcal{G}}$. The process from node j's point of view is illustrated in Figure 1(a).

We define strategies over the entire network by aggregating local rules: A first-stage communication and second-stage estimation strategy pair $\gamma = (\mu, \nu)$ is defined as $\mu = (\mu_1, \mu_2, ..., \mu_N)$ and $\nu = (\nu_1, \nu_2, ..., \nu_N)$, respectively. We refer to $\gamma = (\gamma_1, \gamma_2, ..., \gamma_N)$ as a two-stage strategy. The space of two-stage strategies over $\mathcal{G}$ is given by $\Gamma^{\mathcal{G}} = \bigotimes_{v \in \mathcal{V}} \Gamma_v^{\mathcal{G}}$. It can be seen that $\Gamma^{\mathcal{G}} = \{\gamma \mid \gamma : \mathcal{Y} \to \mathcal{X} \times \mathcal{U}\}$. Here, $\gamma \in \Gamma^{\mathcal{G}}$ is restricted to the strategies which produce $u \in \mathcal{U}$ in accordance with the network $\mathcal{G}$. Consider the set of strategies $\gamma : \mathcal{Y} \to \mathcal{X} \times \mathcal{U}$ which do not take u into account. For example, the centralized estimator which operates over the joint posterior is such a strategy. If we denote the set of u unrestricted strategies by Γ, then, $\Gamma^{\mathcal{G}} \subset \Gamma$. The global view of the strategy is illustrated in Figure 1(b).

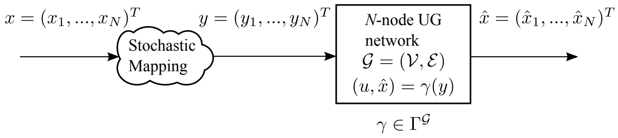
The networked constrained online processing model above provides an abstraction of the subtleties related to the physical, network and other lower layers of the communication architecture. There has been a considerable amount of work on networking sensors including connectivity control [23], Medium

[4]In particular, [18] introduces an additional variable z_j as the channel output to node j. This variable can be treated as a function of the messages sent from the neighbors $ne(j)$ and characterised by a conditional distribution $p(z_j|\overleftarrow{u}_j)$. Examples in which this distribution is specified for modeling binary erasure channels and broadcast channels with interference can be found in [19].

[5]Note that a variety of transmission schemes can be represented by μ_j such as "broadcast" and "peer-to-peer". In order to model the former, $\overrightarrow{\mathcal{U}}_j$ can be replaced with its subset which contains identical messages for all neighbors. Our setting falls into the peer-to-peer type communication in this perspective.



(a)



(b)

Figure 1: Two-stage in-network online processing strategy over an UG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: (a) The viewpoint of node j in $\mathcal{G}$ which evaluates its first-stage communication rule μ_j based on its measurement y_j. In the second-stage, ν_j is evaluated at the incoming messages $\overleftarrow{u}_j$ and y_j and an estimate $\hat{x}_j$ is produced. (b) The global view of the two-stage strategy over $\mathcal{G}$ where a random vector X takes the value x as the outcome of an experiment and induces observations y.

Access Control [24] and multi-hop routing protocols enabling transmission between any two nodes (see, e.g., [23][25],[26]). Therefore, a higher level architecture underpinning the two-stage strategy can be designed using an adequate combination of these results in consideration of the application specific requirements [27, 28]. For the cases that the transmission errors and packet losses cannot be ignored, channel models characterizing these possibilities can be used in the online model as discussed previously.

2.2. *Design problem in a constrained optimization setting*

Given an arbitrary UG $\mathcal{G}$, the selection of a two-stage strategy from $\Gamma^{\mathcal{G}}$ is based on a Bayesian risk function $J(\gamma)$ where $\gamma = (\mu, \nu) \in \Gamma^{\mathcal{G}}$, is constructed as follows: One can select a cost c such that an estimation error penalty for the pair $(x, \hat{x})$ and a cost due to the corresponding set of messages in the network u are assigned, i.e., $c : \mathcal{U} \times \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. For an arbitrary strategy $\gamma \in \Gamma^{\mathcal{G}}$, the corresponding Bayesian risk is given by

$$J(\gamma) \triangleq E\left\{c(U, X, \hat{X}); \gamma\right\} = E\left\{E\left\{c\left(\mu(Y), X, \nu(Y, \mu(Y))\right)|Y\right\}\right\}. \tag{1}$$

Selection of the best two-stage strategy for estimation under communication constraints is, hence, equivalent to solving the constrained optimization problem given by

$$(\text{P}): \qquad \min J(\gamma) \tag{2}$$
$$\text{subject to } \gamma \in \Gamma^{\mathcal{G}}$$

The distribution underlying the expectation in (1) is specified by γ through the density $p(u, \hat{x}|y; \gamma)$ and the equation

$$p(u, \hat{x}, x; \gamma) = \int_{\mathcal{Y}} \mathrm{d}y \, p(u, \hat{x}|y; \gamma) p(y, x), \tag{3}$$

which can be shown after realizing that the tuple $(U, \hat{X}) = \gamma(Y)$ is a random vector conditionally independent of X given Y (denoted by $(U, \hat{X}) \perp\!\!\!\perp X \,|\, Y$) provided that $\gamma = (\gamma_1, ..., \gamma_N) \in \Gamma^{\mathcal{G}}$ is known. Then, the density $p(u, \hat{x}|y)$ is specified by γ and denoted by $p(u, \hat{x}|y; \gamma)$.

Let us consider how local communication and computation rules take part in this density: Once the local rule pair $\gamma_j = (\mu_j, \nu_j)$ is fixed, the conditional density of the outcomes $p(\overrightarrow{u}_j, \hat{x}_j|y_j, \overleftarrow{u}_j; \gamma_j)$ becomes specified. By the two stage mechanism, this density decomposes further as

$$p(\overrightarrow{u}_j, \hat{x}_j|y_j, \overleftarrow{u}_j; \gamma_j) = p(\overrightarrow{u}_j|y_j; \mu_j) p(\hat{x}_j|y_j, \overleftarrow{u}_j; \nu_j).$$

The distribution $p(u, \hat{x}|y; \gamma)$, then, builds upon the local rule pairs following the causal processing provided by γ and the following factorization holds:

$$p(u, \hat{x}|y; \gamma) = \prod_{j \in \mathcal{V}} p(\overrightarrow{u}_j|y_j; \mu_j) p(\hat{x}_j|y_j, \overleftarrow{u}_j; \nu_j). \tag{4}$$

In Problem (P), it can be shown that if there exists an optimal strategy, then there exists an optimal deterministic strategy [29]. Therefore it suffices to consider the deterministic local rule spaces for which case the local first and second stage rules specify the densities involved in Eq.(4) as follows:

$$p(\overrightarrow{u}_j|y_j; \mu_j) = \delta_{\mu_j(y_j)}(\overrightarrow{u}_j) \tag{5}$$

$$p(\hat{x}_j|y_j, \overleftarrow{u}_j; \nu_j) = \delta(\hat{x}_j - \nu_j(y_j, \overleftarrow{u}_j)) \tag{6}$$

where $\delta_m(n)$ is the Kronecker delta and δ is the Dirac delta distribution. After substituting Eq.s (5) and (6) into Eq.(4) and Eq.(3), the distribution underlying the Bayesian risk is specified.

We provide a table of symbols introduced in this section in Table 1 for helping the reader throughout the rest of the article.

3. Team Decision Theoretic Formulation

Problem (P) in (2) is a typical team decision problem [30]. It is often not possible to find solutions with global optimality guarantees(see, e.g., [29]). A convenient solution approach which has been used

Algorithm 1 Iterations converging to a person-by-person optimal strategy.

```
Choose γ^0 = (γ_1^0, γ_2^0, ..., γ_N^0) ∈ Γ^G and ε ∈ ℝ^+          ▷ Initialize
l ← 0
repeat
    l ← l + 1
    for j = N, N − 1, . . . , 1 do
        γ_j^l = arg min_{γ_j ∈ Γ_j^G} J(γ_1^{l−1}, ..., γ_{j−1}^{l−1}, γ_j, γ_{j+1}^l, ..., γ_N^l)          ▷ Update
    end for
until J(γ^{l−1}) − J(γ^l) < ε          ▷ Check
```

in a variety of similar contexts including quantizer design for minimum distortion [31, 32] and distributed estimation [33, 34] is to use necessary (but not sufficient) conditions of optimality to achieve nonlinear Gauss-Seidel iterations converging to a person-by-person (pbp) optimal strategy [29][18]: At the pbp optimal point $\gamma^* \in \Gamma^{\mathcal{G}}$, it holds that $J(\gamma_j^*, \gamma_{\backslash j}^*) \leq J(\gamma_j, \gamma_{\backslash j}^*)$ for all $\gamma_j \in \Gamma_j^{\mathcal{G}}$ where $\backslash j$ denotes $\mathcal{V} \setminus j$ and $\gamma_{\backslash j}^* = \{\gamma_1^*, \gamma_2^*, ..., \gamma_{j-1}^*, \gamma_{j+1}^*, ..., \gamma_N^*\}$[6]. In other words, no improvement to $J(\gamma^*)$ can be obtained by varying only a single local rule γ_j^*. The strategies that satisfy this equilibrium condition are solutions to a relaxation of (P) in which one is interested in finding $\gamma^* = (\gamma_1^*, ..., \gamma_n^*)$ such that

$$\gamma_j^* = \arg \min_{\gamma_j \in \Gamma_j} J(\gamma_j, \gamma_{\backslash j}^*) \tag{7}$$

for all $j \in \{1, 2, ..., N\}$. The strategy γ^* is referred to as a pbp optimal strategy. The iterations given by Algorithm 1 converge to such a solution starting with an arbitrary set of local rules.

It is useful to note that the converged strategy depends on the initialization, in general. Therefore, it is a good practice to start the iterations with a reasonable selection of initial rules and use Algorithm 1 to improve upon them. For the example scenarios presented in Section 5, the iterative approach delivers a consistent performance with different initializations.

For the *detection* problem, an extensive study of pbp optimal solutions for a number of strategy classes can be found in [18]. One of these classes exhibits directed acyclic communication and computation structures and can equivalently be represented by DAGs [19]. It has been shown that in the case of two-stage strategies over undirected communication topologies, pbp optimal set of local rules lie in a finitely parameterized subspace of $\Gamma^{\mathcal{G}}$, and hence errors involved in their computation is mainly due to finite machine

[6] When it is clear from the context, we denote $\{x_i \mid i \in I\}$ by x_I where I is an index set for the collection of variables $\{x_1, x_2, ..., x_N\}$.

precision. This is partly because X_js of a detection problem, contrary to the estimation setting, take values from finite sets. The communication and computation structure of a two-stage strategy can equivalently be represented through a bipartite graph (Chp. 4 of [18]). Such graphs are directed and acyclic structures and, hence, two-stage rules can be investigated using the results for the detection problem over a DAG (provided that certain assumptions hold).

In our estimation setting over an undirected graph, we follow a similar approach and exploit the pbp optimality condition for decentralized estimation strategies over DAGs [20][7][8].

We start by unwrapping the communication and computation structure of two-stage strategies over undirected communication topologies onto directed acyclic bipartite graphs. The two-stage operation enables us to represent the same platform with two nodes of different types. The nodes of the bipartite graph $\mathcal{B} = ((\mathcal{V}, \mathcal{V}'), \mathcal{F})$ are identified by considering the set of nodes in the undirected graph $\mathcal{G}$, i.e., $\mathcal{V}$, and its replicate $\mathcal{V}' \triangleq \{j' \mid j \in \mathcal{V}\}$ as a pair and assigning the communication rules and the estimation rules to $\mathcal{V}$ and $\mathcal{V}'$, respectively. The edges of the bipartite graph connect communication nodes in $\mathcal{V}$ to the estimation rules of the neighbor nodes in $\mathcal{V}'$. In other words, $(j, i') \in \mathcal{F}$ if $i \in ne(j)$ in $\mathcal{G}$. For example, consider the undirected communication topology given in Figure 2(a). The two-stage strategy over this UG is explicitly shown in Figure 2(b). The unwrapped directed acyclic communication and computation structure of the two-stage strategy which is a bipartite graph is shown in Figure 2(c). Nodes $1 - 4$ in $\mathcal{V}$ perform only the communication rules, i.e., μ_js. Likewise, nodes $1' - 4'$ in $\mathcal{V}'$ are associated only with the estimation rules, i.e., ν_js. Node j and j' correspond to the same physical platform but different processing tasks, in this respect.

At this point, it is useful to contrast the two-stage strategy design problem with that for an FC estimator in a star-topology [33]. In the conventional setting, the design goal is to find an estimation rule for the FC and quantizers for the peripheral sensors which minimize the expected cost of estimation errors. The FC receives messages from all of the other sensors, however, communication is not penalized. The two-stage strategy we consider decentralizes the estimation task in a way that each node can be viewed as a local FC

[7] In principle, it is possible to obtain the estimation results presented in this section starting from the detection results in [18] and performing the marginalizations in the variables X_js and $\hat{X}_j$s through appropriate integrations (as opposed to summations) under error-free and "peer-to-peer" transmission assumptions. In part because $\mathcal{X}_j$s are nondenumerable, our problem, contrary to the detection setting, does not lead to pbp optimal local rules that can be characterized with a finite set of parameters, in general.

[8] In the case of a dynamic problem in which $p(x)$ varies over time, the strategies can be updated accordingly. Investigation of efficient methods for updating strategies in dynamic problems is left beyond the scope of this work.

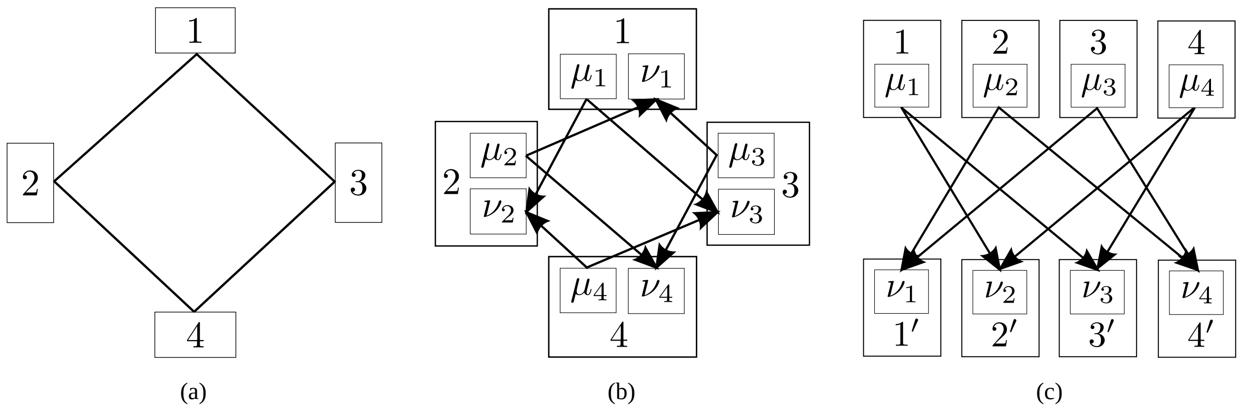


Figure 2: (a) A loopy UG of 4 nodes. (b) The two stage strategy over the UG. (c) The bi-partite DAG counterpart of the two-stage online processing: Nodes 1–4 correspond to platforms 1–4 but only performing the communication rules, whereas nodes 1′–4′ correspond to platforms 1–4 but only performing the estimation rules.

with its neighbors as peripherals (e.g., the estimation nodes $1' - 4'$ in Figure 2(c) can be viewed as FCs of their local networks) and the communication rules are not restricted to quantizers. These star networks are *coupled* in the two-stage strategy design as all the estimation and communication rules that constitute the strategy are considered jointly through the cost function $c(\hat{x}, x, u)$.

Next, we make a set of assumptions:

Assumption 1. *The global cost function is the sum of costs due to the communication rules and the decision rules, which are in turn additive over the nodes:*

$$\begin{aligned} c(u, \hat{x}, x) &= c^d(\hat{x}, x) + \lambda c^c(u, x) \\ c^d(\hat{x}, x) &= \sum_{i \in \mathcal{V}} c_i^d(\hat{x}_i, x_i) \\ c^c(u, x) &= \sum_{i \in \mathcal{V}} c_i^c(\overrightarrow{u}_i, x) \end{aligned} \tag{8}$$

Here, λ appears as a unit conversion constant and can be interpreted as the equivalent estimation penalty per unit communication cost [18]. Hence $J(\gamma) = J_d(\gamma) + \lambda J_c(\gamma)$ where $J_d(\gamma) = E\{c^d(\hat{x}, x); \gamma\}$ and $J_c(\gamma) = E\{c^c(u, x); \gamma\}$ respectively[9].

[9]Note that convex combinations of dual objectives, i.e., $J'(\gamma) = \alpha J_d(\gamma) + (1-\alpha) J_c(\gamma)$, yield pareto-optimal curves parameterized by α. This setting preserves the pareto-optimal front since $\lambda = (1-\alpha)/\alpha$ and $J(\gamma) \propto J'(\gamma)$ yielding a graceful degradation of the estimation performance as λ is increased.

Assumption 2. *(Conditional Independence) The noise processes of the sensors are mutually independent and hence given the state of X, the observations are conditionally independent, i.e.,* $p(x, y) = p(x) \prod_{i=1}^{N} p(y_i|x)$.

Assumption 3. *(Measurement Locality) Every node j observes y_j due to only x_j, i.e.,* $p(y_j|x) = p(y_j|x_j)$.

Under these conditions, it is possible to apply *Corollary 3.4* in [20], which reveals the structure of the pbp optimal local communication and estimation rules in strategies over DAGs, to the bipartite representation of the two-stage strategies. Before stating this result, let us define two-step neighbors of j by $ne^2(j) \triangleq \cup_{i \in ne(j)} ne(i) \setminus j$.

Proposition 3.1. *(Adaptation of Proposition 4.3 in [18] for estimation) Suppose that Assumptions 1-3 hold and suppose we are given a pbp optimal two-stage strategy $\gamma^* = (\gamma_1^*, \ldots \gamma_N^*)$ over an undirected graph. If all the local rules other than the j^{th} are fixed at the optimum point, the j^{th} optimal rule can be characterized as follows: The communication rule (evaluated at stage-one) is given by*

$$\mu_j^*(y_j) = \arg \min_{\overrightarrow{u}_j \in \overrightarrow{\mathcal{U}}_j} \int_{\mathcal{X}_j} \mathrm{d}x_j \, p(y_j|x_j) \alpha_j(\overrightarrow{u}_j, x_j; \nu_{ne(j)}^*, \mu_{ne^2(j)}^*) \tag{9}$$

for all $y_j \in \mathcal{Y}_j$ with nonzero probability, where

$$\alpha_j(\overrightarrow{u}_j, x_j; \nu_{ne(j)}^*, \mu_{ne^2(j)}^*) \propto p(x_j)[\lambda c_j^c(\overrightarrow{u}_j, x_j) + C_j(\overrightarrow{u}_j, x_j; \nu_{ne(j)}^*, \mu_{ne^2(j)}^*)]. \tag{10}$$

The estimation rule (evaluated at stage-two) is given by

$$\nu_j^*(y_j, \overleftarrow{u}_j) = \arg \min_{\hat{x}_j \in \mathcal{X}_j} \int_{\mathcal{X}_j} \mathrm{d}x_j \, p(y_j|x_j) \beta_j(x_j, \hat{x}_j, \overleftarrow{u}_j; \mu_{ne(j)}^*) \tag{11}$$

for all $y_j \in \mathcal{Y}_j$ and for all $\overleftarrow{u}_j \in \overleftarrow{\mathcal{U}}_j$ with nonzero probability where

$$\beta_j(x_j, \hat{x}_j, \overleftarrow{u}_j; \mu_{ne(j)}^*) \propto p(x_j) P_j(\overleftarrow{u}_j|x_j; \mu_{ne(j)}^*) c_j^d(\hat{x}_j, x_j). \tag{12}$$

The term $P_j(\overleftarrow{u}_j|x_j; \mu_{ne(j)}^*)$ in Eq.(12) is the (incoming) message likelihood and given by

$$P_j(\overleftarrow{u}_j|x_j; \mu_{ne(j)}^*) = \int_{\mathcal{X}_{ne(j)}} \mathrm{d}x_{ne(j)} \, p(x_{ne(j)}|x_j) \prod_{i \in ne(j)} P_{i \to j}(u_{i \to j}|x_i; \mu_{ne(j)}^*) \tag{13}$$

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

$$P_{i \to j}(u_{i \to j}|x_i; \mu_i^*) = \sum_{\overrightarrow{u}_i \setminus u_{i \to j}} p(\overrightarrow{u}_i|x_i; \mu_i^*) \tag{14}$$

for all $u_{i\to j} \in \mathcal{U}_{i\to j}$ *where*

$$p(\overrightarrow{u}_i|x_i;\mu_i^*) = \int_{\mathcal{Y}_i} \mathrm{d}y_i\, p(y_i|x_i)p(\overrightarrow{u}_i|y_i;\mu_i^*). \tag{15}$$

The term $C_j(\overrightarrow{u}_j, x_j; \nu^*_{ne(j)}, \mu^*_{ne^2(j)})$ *in Eq.*(10) *is the total expected cost and given by*

$$C_j(\overrightarrow{u}_j, x_j; \nu^*_{ne(j)}, \mu^*_{ne^2(j)}) = \sum_{i\in ne(j)} C_{i\to j}(u_{j\to i}, x_j; \nu_i^*, \mu^*_{ne(i)}) \tag{16}$$

for all $\overrightarrow{u}_j \in \overrightarrow{\mathcal{U}}_j$ *with terms capturing the influence of* j *on* $i \in ne(j)$ *given by*

$$C_{i\to j}(u_{j\to i}, x_j; \nu_i^*, \mu^*_{ne(i)}) = \int_{\mathcal{X}_{ne(i)\setminus j}} dx_{ne(i)\setminus j} \int_{\mathcal{X}_i} dx_i\, p(x_{ne(i)\setminus j}, x_i|x_j)$$
$$\times \sum_{u_{ne(i)\setminus j}} \prod_{j'\in ne(i)\setminus j} P_{j'\to i}(u_{j'\to i}|x_{j'};\mu^*_{j'}) I_i(\overleftarrow{u}_i, x_i; \nu_i^*) \tag{17}$$

such that

$$I_i(\overleftarrow{u}_i, x_i; \nu_i^*) = \int_{\mathcal{Y}_i} \mathrm{d}y_i \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i\, c_i^d(\hat{x}_i, x_i) p(\hat{x}_i|y_i, \overleftarrow{u}_i; \nu_i^*) p(y_i|x_i). \tag{18}$$

Proof. As discussed at the beginning of this section, two-stage strategies over undirected graphs can equivalently be represented by strategies over DAGs. Under Assumptions 1–2, *Corollary 3.4* in [20] is valid over the bipartite directed acyclic model associated with the two-stage strategies over the undirected graph $\mathcal{G}$. Consider the bipartite DAG $\mathcal{B} = ((\mathcal{V}, \mathcal{V}'), \mathcal{F})$ associated with the undirected graph $\mathcal{G}$. Proposition 3.1 is obtained after applying *Corollary 3.4* in [20] on $\mathcal{B}$ and then refolding it back to $\mathcal{G}$ by substituting j for all $j' \in \mathcal{V}'$. □

Proposition 3.1 provides a variational characterization of the j^{th} communication and estimation rules, given a pbp optimal two-stage strategy[10]. Let us use a simpler notation for the terms on the left hand side (LHS) of Eq.s(13) and (16) and denote them by $P_j(\overleftarrow{u}_j|x_j)$ and $C_j(\overrightarrow{u}_j, x_j)$, respectively. Considering Eq.s(13) and (14), $P_j(\overleftarrow{u}_j|x_j)$ is a likelihood function for x_j inducing $\overleftarrow{u}_j$. Eq.s(16)-(18) reveal that $C_j(\overrightarrow{u}_j, x_j)$ is the total expected cost induced on the neighbors by transmitting $\overrightarrow{u}_j$, i.e., $E\{c^d(\hat{x}_{ne(j)}, x_{ne(j)})|\overrightarrow{u}_j, x_j; \nu^*_{ne(j)}, \mu^*_{ne^2(j)}\}$. Since $p(x_j)p(y_j|x_j)P(\overleftarrow{u}_j|x_j) \propto p(x_j|y_j, \overleftarrow{u}_j)$ holds under Assumptions 2-3, the j^{th} optimal communication rule selects the message that results with a minimum contribution to the overall cost and the optimal estimation rule selects $\hat{x}_j$ that yields the minimum expected penalty given y_j and $\overleftarrow{u}_j$. For example, if

[10] The integrals over $\mathcal{X}_j$ and $\mathcal{Y}_j$ should be interpreted in accordance with the dimensionality of their domains.

$c^d_j(\hat{x}_j, x_j) = (\hat{x}_j - x_j)^2$ as in the conventional mean squared error (MSE) estimator, then the estimation rule in Eq.(11) can be expressed in closed form as

$$\hat{x}_j = \nu^*_j(y_j, \overleftarrow{u}_j) = \frac{\int_{\mathcal{X}_j} \mathrm{d}x_j\, x_j p(x_j) p(y_j|x_j) P_j(\overleftarrow{u}_j|x_j)}{\int_{\mathcal{X}_j} \mathrm{d}x_j\, p(x_j) p(y_j|x_j) P_j(\overleftarrow{u}_j|x_j)}. \quad (19)$$

Since $P_j(\overleftarrow{u}_j|x_j) = p(\overleftarrow{u}_j|x_j; \mu^*_{ne(j)})$ is the likelihood of the incoming messages and the conditional independence relation $\overleftarrow{U}_j \perp\!\!\!\perp Y_j \,|\, X_j$ holds, then

$$p(x_j, y_j, \overleftarrow{u}_j) = p(x_j) p(y_j|x_j) p(\overleftarrow{u}_j|x_j)$$

and the denominator in Eq.(19) is nothing but $p(y_j, \overleftarrow{u}_j) = p(y_j, \overleftarrow{u}_j; \mu^*_{ne(j)})$. Consequently, the local estimation rule is the expected value of the posterior given the local measurement and incoming messages given by

$$\hat{x}_j = \nu^*_j(y_j, \overleftarrow{u}_j) = \int_{\mathcal{X}_j} \mathrm{d}x_j\, x_j p(x_j | y_j, \overleftarrow{u}_j; \mu^*_{ne(j)}).$$

Based on Proposition 3.1, it is possible to tailor the `Update` step of Algorithm 1 to obtain an iterative scheme for finding a pbp optimal two-stage strategy. The treatment of the terms in Eq.s(10), (12)-(18) as operators that can act on any set of local rules, not necessarily optimal, results with Algorithm 2. Note that, these steps can be carried out in a message passing fashion. In the first pass (`Update Step 1`), all nodes compute and send node-to-node likelihood terms to their neighbors. In the second pass (`Update Step 2`), upon reception of these messages, all nodes update their (incoming) message likelihoods and estimation rules. Then, they compute and send expected cost messages to their neighbors. After receiving cost messages from neighbors, each node updates its communication rule (`Update Step 3`). Owing to the message passing structure, the complexity of optimization is bounded by the node with the highest degree rather than the number of nodes. Such a structure is also advantageous in the case of a network self-organization requirement.

Finally, the value of the Bayesian risk function at the l^{th} iteration is easily found in terms of the expressions discussed above as

$$J(\gamma^l) = \sum_{i \in \mathcal{V}} G^d_i(\nu^l_i) + \lambda \sum_{i \in \mathcal{V}} G^c_i(\mu^l_i), \quad (20)$$

where the per node costs are given by

$$G^d_i(\nu^l_i) = \sum_{\overleftarrow{u}_i} \int_{\mathcal{X}_i} \mathrm{d}x_i p(x_i) P^{l+1}_i(\overleftarrow{u}_i|x_i) I_i(\overleftarrow{u}_i, x_i; \nu^l_i), \quad (21)$$

$$G^c_i(\mu^l_i) = \sum_{\overrightarrow{u}_i} \int_{\mathcal{X}_i} \mathrm{d}x_i c^c_i(\overrightarrow{u}_i, x_i) p(x_i) p(\overrightarrow{u}_i|x_i; \mu^l_i). \quad (22)$$

Algorithm 2 Iterations converging to a pbp optimal two-stage strategy over a UG $\mathcal{G}$.

Choose $\gamma^0 = (\gamma_1^0, \gamma_2^0, \ldots, \gamma_N^0) \in \Gamma^{\mathcal{G}}$ and $\varepsilon \in \mathbb{R}^+$ ▷ `Initialize`

$l \leftarrow 0$

repeat

  $l \leftarrow l + 1$

  for $i = 1, 2, \ldots, N$ **do** ▷ `(Update Step 1)`

    Find the node-to-node likelihood messages $P^l_{i\to j} = P_{i\to j}(u_{i\to j}|x_i; \mu_i^{l-1})$ for $j \in ne(i)$ using Eq.s(15) and (14).

  end for

  for $j = 1, 2, \ldots, N$ **do** ▷ `(Update Step 2)`

    Find the incoming message likelihood ${P_j}^l$ by substituting $P^l_{i\to j}$s into Eq. (13).

    Find the estimation rule ν^l_j by substituting ${P_j}^l$ in Eq.s(12) and (11).

    Find the cost messages $C^l_{j\to i}$ for $i \in ne(j)$ by using ν^l_j and $P^l_{i\to j}$ in Eq.s(18) and (17).

  end for

  for $j = 1, 2, \ldots, N$ **do** ▷ `(Update Step 3)`

    Find the communication rule μ^l_j by substituting $C^l_{i\to j}$ from $i \in ne(j)$ into Eq.s(16),(10) and (9).

  end for

until $J(\gamma^{l-1}) - J(\gamma^l) < \varepsilon$ ▷ `Check`

4. MC Optimization Framework for two-stage in-network processing strategies over UGs

In this Section, we develop Monte Carlo (MC) methods to realize Algorithm 2 introduced in Section 3. Algorithm 2 results with a pbp optimal processing strategy whose structure is captured by the operators in Proposition 3.1. It is not possible to evaluate these operators for arbitrary selections of, e.g., priors $p(x_j)$s, likelihoods $p(y_j|x_j)$s or $\gamma_{\setminus j} \in \Gamma^{\mathcal{G}}_{\setminus j}$, in general. Instead, we consider a fixed set of particles at each node and approximate the aforementioned operators using MC methods such as Importance Sampling (IS) [35, 36]. The resulting algorithm which is detailed in this section carries out strategy optimization through passing messages represented by weighted particles[11].

We use IS with independent samples generated from two proposal distributions $s_j(x_j)$ and $q_j(y_j)$ over

[11]Similar decentralized algorithms based on transmissions of weighted particles include particle Belief Propagation algorithms (see, e.g., [37, 38]) for estimation.

$\mathcal{X}_j$ and $\mathcal{Y}_j$, respectively for node j:

$$S_j \triangleq \{x_j^{(1)}, x_j^{(2)}, ..., x_j^{(M_j)}\} \text{ such that } x_j^{(m)} \sim s_j(x_j) \text{ for } m = 1, 2, ..., M_j, \quad (23)$$

and,

$$Q_j \triangleq \{y_j^{(1)}, y_j^{(2)}, ..., y_j^{(P_j)}\} \text{ such that } y_j^{(p)} \sim q_j(y_j) \text{ for } p = 1, 2, ..., P_j. \quad (24)$$

These proposal distributions can be selected as the local marginals $p(x_j)$ and $p(y_j)$. This sampling strategy has been previously used in similar message passing algorithms (see, for example, [38] and the references therein). Use of heavy tailed distributions would improve the small sample size variance of IS [36]. Although the sizes of S_j and Q_j might vary, we assume that $M_j = M$ and $P_j = P$ for $j \in \mathcal{V}$ for the simplicity of the discussion throughout.

We fix these particle sets in order to reduce the communication load of the optimization by not having to transmit particles at every iteration but transmit them only once and communicate the weights for the rest of the iterations. This approach is similar to that proposed in [38] for particle BP algorithms, and, has also been used in [20] for optimizing decentralized strategies over DAGs.

Using these sample sets, we make successive approximations to the expressions constituting the j^{th} pbp optimal local rule given in Proposition 3.1. First, we approximate to the local rule pair in Section 4.1. Then, we apply the IS rule to the incoming message likelihood (Sec. 4.2). In Section 4.3, we tackle computations regarding the expected cost term. Finally, in Section 4.4, we employ all the previous steps simultaneously in Algorithm 2 and obtain a Monte Carlo optimization scheme such that the message passing structure is preserved.

4.1. Approximating the person-by-person optimal local rule

Let us consider Proposition 3.1 for the variational form of the j^{th} communication and estimation rules in the case of an arbitrary $\gamma_{\backslash j}$ not necessarily optimal. We approximate Eq.s(9) and (11) since it is often not possible to compute these integrals, exactly, for arbitrary selections of the factors that construct α_j and β_j (given in Eq.s(10) and (12), respectively).

We simplify our notation by hiding the dependence of the operators in Proposition 3.1 to the local rules in $\gamma_{\backslash j}$. For example, we denote the incoming message likelihood in Eq.(13) and the total expected cost in Eq.(16) by $P_j(\overleftarrow{u}_j|x_j)$ and $C_j(\overrightarrow{u}_j, x_j)$, respectively, where the underlying rules are obvious from the context.

We use the sample set S_j in Eq.(23) for finding an IS approximation to the communication rule in Eq.(9) and obtain

$$\mu_j(y_j) \approx \arg\min_{\overrightarrow{u}_j\in\overrightarrow{\mathcal{U}}_j} \frac{1}{\sum_{m'=1}^{M}\omega_j^{(m')}} \sum_{m=1}^{M} \omega_j^{(m)} p(y_j|x_j^{(m)})[\lambda c_j^c(\overrightarrow{u}_j, x_j^{(m)}) + C_j(\overrightarrow{u}_j, x_j^{(m)})], \tag{25}$$

$$\omega_j^m = p(x_j^{(m)})/s_j(x_j^{(m)}), \tag{26}$$

for all $y_j \in \mathcal{Y}_j$ with non-zero probability.

For the local estimation rule given in (11), a similar approximation is given by

$$\nu_j(y_j, \overleftarrow{u}_j) \approx \arg\min_{\hat{x}_j\in\mathcal{X}_j} \frac{1}{\sum_{m'=1}^{M}\omega_j^{(m')}} \sum_{m=1}^{M} \omega_j^m p(y_j|x_j^{(m)}) P_j(\overleftarrow{u}_j|x_j^{(m)}) c_j^d(\hat{x}_j, x_j^{(m)}), \tag{27}$$

for all $y_j \in \mathcal{Y}_j$ and $\overleftarrow{u}_j \in \overleftarrow{\mathcal{U}}_j$ with non-zero probability, using the IS weights in Eq.(26).

Example 4.1. *Consider the squared error penalty for the estimation error, i.e.,* $c_j^d(\hat{x}_j, x_j) = (\hat{x}_j - x_j)^2$. *Then the pbp optimal estimation rule local to node* j *as given in the variational form by Eq.*(27) *yields*

$$\nu_j(y_j, \overleftarrow{u}_j) \approx \frac{\sum_{m=1}^{M} \omega_j^{(m)} x_j^{(m)} p(y_j|x_j^{(m)}) P_j(\overleftarrow{u}_j|x_j^{(m)})}{\sum_{m=1}^{M} \omega_j^{(m)} p(y_j|x_j^{(m)}) P_j(\overleftarrow{u}_j|x_j^{(m)})}.$$

4.2. Approximating the message likelihood function

We consider the message likelihood function $P_j(\overleftarrow{u}_j|x_j)$ in the right hand side of (27) given by Eq.(13) together with the recursion involving Eq.s(14) and (15). We find an IS approximation for evaluations of $P_j(\overleftarrow{u}_j|x_j)$ at $x_j \in S_j$ and $\overleftarrow{u}_j \in \overleftarrow{\mathcal{U}}_j$ as follows: We first consider $p(\overrightarrow{u}_i|x_i;\mu_i)$ in (15). We use the IS rule with the sample set Q_j generated from the local proposal density $q_i(y_i)$:

$$\tilde{p}(\overrightarrow{u}_i|x_i^{(m)};\mu_i) \triangleq \frac{1}{\sum_{p=1}^{P}\omega_i^{(m)(p)}} \sum_{p=1}^{P} \omega_i^{(m)(p)} \delta_{\mu_i(y_i^{(p)})}(\overrightarrow{u}_i) \tag{28}$$

$$\omega_i^{(m)(p)} = \frac{p(y_i^{(p)}|x_i^{(m)})}{q_i(y_i^{(p)})}$$

for $\overrightarrow{u}_i \in \mathcal{U}_i$ and $x_i^{(m)} \in S_i$.

Note that the node-to-node likelihood $P_{i\to j}$ in (14) is a marginalization of $p(\overrightarrow{u}_i|x_i;\mu_i)$ and can be estimated by substituting $\tilde{p}$ in (14). Let us denote this term by $\tilde{P}_{i\to j}$.

Second, we consider $P_j(\overleftarrow{u}_j|x_j)$ in (13) and construct a sample set at node j by using the particle sets S_is local to the neighbors. The m^{th} element in this set is a vector obtained by concatenating the m^{th} elements from S_is, i.e., we construct $S_{ne(j)} \triangleq \{x^{(m)}_{ne(j)} | x^{(m)}_{ne(j)} = (x^{(m)}_i)_{i\in ne(j)}\}$. Note that these points are generated from the product of proposals, i.e., $x^{(m)}_{ne(j)} \sim \prod_{i\in ne(j)} s_i(x_i)$. We consider using this sample set with the IS method and equivalently the proposal density $\prod_{i\in ne(j)} s_i(x_i)$. Then, the integral in the RHS of Eq.(13) can be approximated with

$$\tilde{P}_j(\overleftarrow{u}_j | x^{(m)}_j) \triangleq \frac{1}{\sum_{m'=1}^{M} \omega^{(m)(m')}_j} \sum_{m'=1}^{M} \omega^{(m)(m')}_j \prod_{i\in ne(j)} \tilde{P}_{i\to j}(u_{i\to j} | x^{(m')}_i), \tag{29}$$

$$\omega^{(m)(m')}_j = \frac{p(x^{(m')}_{ne(j)} | x^{(m)}_j)}{\prod_{i\in ne(j)} s_i(x^{(m')}_i)}.$$

We replace the P_j term in the RHS of Eq.(27) by $\tilde{P}_j$ and obtain an approximately pbp optimal estimation rule through these successive IS approximations.

4.3. Approximating the expected cost term

We consider the expected cost term C_j in the RHS of the communication rule approximation in (25). This term is given by Eq.s(16)–(18) and we begin with approximating to the conditional estimation risk $I_i(\overleftarrow{u}_i, x_i; \nu_i)$. After substituting from (6) into (18), we obtain

$$I_i(\overleftarrow{u}_i, x_i; \nu_i) = \int_{\mathcal{Y}_i} \mathrm{d}y_i c^d_i(\nu_i(y_i, \overleftarrow{u}_i), x_i) p(y_i|x_i).$$

For the RHS of the expression above, we use $q_i(x_i)$ as the proposal distribution of the IS rule and utilize the sample set Q_i (Eq.(24)). Then, the conditional expected risk is estimated by

$$\tilde{I}_i(\overleftarrow{u}_i, x^{(m)}_i; \nu_i) \triangleq \frac{1}{\sum_{p=1}^{P} \omega^{(m)(p)}_i} \sum_{p=1}^{P} \omega^{(m)(p)}_i c^d_i(\nu_i(y^{(p)}_i, \overleftarrow{u}_i), x^{(m)}_i) \tag{30}$$

$$\omega^{(m)(p)}_i = \frac{p(y^{(p)}_i | x^{(m)}_i)}{q_i(y^{(p)}_i)}$$

for all $\overleftarrow{u}_i \in \overleftarrow{\mathcal{U}}_i$ and $x^{(m)}_i \in S_i$.

Now, let us consider the approximate evaluation of the node-to-node cost messages $C_{i\to j}$ given by Eq. (17). We employ IS for approximately evaluating the RHS of Eq. (17) at all possible $(u_{j\to i}, x^{(m)}_j)$ pairs such that $u_{j\to i} \in \mathcal{U}_{j\to i}$ and $x^{(m)}_j \in S_j$. Similar to the discussion on approximating the message likelihood term, we consider a sample set constructed by concatenating the m^{th} elements from the usual sets local to neighbors of i other than j, i.e.,

$$S_{x_{ne(i)\setminus j}} \triangleq \{x^{(m)}_{ne(i)\setminus j} | x^{(m)}_{ne(i)\setminus j} = (x^{(m)}_{j'})_{j'\in ne(i)\setminus j}\}$$

This set can equivalently be treated as points generated from $\prod_{j' \in ne(i)\setminus j} s_{j'}(x_{j'})$. Together with S_i, we use the IS approximation to RHS of Eq.(17) and obtain

$$\tilde{C}_{i\to j}(u_{j\to i}, x_j^{(m)}) \triangleq \sum_{u_{ne(i)\setminus j}} \frac{1}{\sum_{m'=1}^{M} \omega_i^{(m)(m')}} \sum_{m'=1}^{M} \omega_i^{(m)(m')} \prod_{j' \in ne(i)\setminus j} \tilde{P}_{j'\to i}(u_{j'\to i}|x_{j'}^{(m')}) \tilde{I}_i(\overleftarrow{u}_i, x_i^{(m')}; \nu_i), \quad (31)$$

$$\omega_i^{(m)(m')} = \frac{p(x_{ne(i)\setminus j}^{(m')}, x_i^{(m')}|x_j^{(m)})}{p(x_i^{(m')}) \prod_{j' \in ne(i)\setminus j} s_{j'}(x_{j'}^{(m')})}.$$

After replacing $C_{i\to j}$ with $\tilde{C}_{i\to j}$ in the total estimation risk in Eq. (16) and the approximate local communication rule in Eq.(25), a further approximation denoted by $\tilde{\mu}_j$ is obtained.

4.4. MC optimization of two-stage in-network processing strategies over UGs

In Sections 4.1–4.3, based on Proposition 3.1, we provided a Monte Carlo framework for approximating the j^{th} local rule in the pbp optimal form given an arbitrary $\gamma_{\setminus j}$. In particular, we obtained $(\tilde{\mu}_j, \tilde{\nu}_j)$ using the IS rule with proposal distributions which might be selected simply as local marginals.

Once the RHSs of all the expressions in the MC framework are considered as operators, we can approximate all local rules in a strategy simultaneously and plug them into Algorithm 2. The procedure we obtain with this approach is given in Algorithm 3. Note that, the message passing structure of the computations is maintained: Before proceeding with the iterations, the nodes exchange S_is with their neighbors. In the first stage of the iterations, the IS weights of the node-to-node likelihoods are transmitted to the neighbors. It suffices to transmit these sets as arrays of weights for each admissible link symbol since S_is are already known to neighbors. In the second stage of the iterations, the cost messages are exchanged, again, as ordered real arrays for each symbol. The node-to-node likelihood from node i to j is, then, of length $M_i \left|\mathcal{U}_{i\to j}\right|$, whereas that of the cost message is $M_j \left|\mathcal{U}_{i\to j}\right|$. In the examples we present in Section 5, convergence is achieved after only a few iterations.

Finally, the value of the Bayesian risk function corresponding to the strategy at the l^{th} iteration, i.e., $J(\gamma^l) = J_d(\gamma^l) + \lambda J_c(\gamma^l)$ given by Eq.s(20)–(22), can be computed approximately by

$$\tilde{J}(\tilde{\gamma}^l) = \sum_{i \in \mathcal{V}} \tilde{G}_i^d(\tilde{\nu}_i^l) + \lambda \sum_{i \in \mathcal{V}} \tilde{G}_i^c(\tilde{\mu}_i^l) \quad (32)$$

where

$$\tilde{G}_i^d(\tilde{\nu}_i^l) = \sum_{\overleftarrow{u}_i, m} \tilde{P}_i^{l+1}(\overleftarrow{u}_i|x_i^{(m)}) \tilde{I}_i^l(\overleftarrow{u}_i, x_i^{(m)}; \tilde{\nu}_i^l), \quad (33)$$

$$\tilde{G}_i^c(\tilde{\mu}_i^l) = \sum_{\overrightarrow{u}_i, m} c_i^c(\overrightarrow{u}_i, x_i^{(m)}) \tilde{p}(\overrightarrow{u}_i|x_i^{(m)}; \tilde{\mu}_i^l). \quad (34)$$

Algorithm 3 Iterations converging to an approximate pbp optimal two-stage in-network processing strategy over a UG $\mathcal{G}$.

```
Choose γ^0 = (γ_1^0, γ_2^0, ..., γ_N^0) ∈ Γ^G and ε ∈ ℝ^+                ▷ Initialize
l ← 0
repeat
    l ← l + 1
    for i = 1, 2, ..., N do                                            ▷ (Update Step 1)
        Find the node-to-node likelihood messages P̃^l_{i→j} = P̃_{i→j}(u_{i→j}|x_i; μ̃_i^{l−1}) at u_{i→j} ∈ U_{i→j}, x_i ∈ S_i for j ∈ ne(i) using Eq.s(28) and (14).
    end for
    for j = 1, 2, ..., N do                                            ▷ (Update Step 2)
        Find the incoming message likelihood P̃_j^l by substituting P̃^l_{i→j}s into Eq. (29).
        Find the estimation rule ν̃_j^l by substituting P̃_j^l in Eq.(27).
        Find the cost messages C̃^l_{j→i} at u_{i→j} ∈ U_{i→j}, x_j ∈ S_j for i ∈ ne(j) by using ν̃_j^l and P̃^l_{i→j} in Eq.s(30) and (31).
    end for
    for i = 1, 2, ..., N do                                            ▷ (Update Step 3)
        Find the communication rule μ̃_j^l by substituting C̃^l_{i→j}s into Eq.s(16) and (25)
    end for
until τ(J̃(γ̃^l), J̃(γ̃^{l−1}), ..., J̃(γ̃^0)) < ε                        ▷ Check
```

In contrary to $\{J(\gamma^l)\}$, the sequence of approximated objectives, i.e., $\{\tilde{J}(\tilde{\gamma}^l)\}$, is not necessarily non-increasing. Nevertheless, note that the error sequence $err[l] \triangleq J(\gamma^l) - \tilde{J}(\tilde{\gamma}^l)$ will be identically zero with probability one as $M, P \to \infty$. Investigation of an operator τ (`Check` step of Algorithm 3) that would yield a non-increasing error sequence with high probability for finite M, P could be a topic for future work.

5. Examples

In this section, we demonstrate our MC-based decentralized estimation framework in various scenarios including Gaussian priors, non-Gaussian priors, and large random graphs. We use local marginals as IS proposal distributions and compare the performances of the optimized strategies with those of the centralized and the myopic estimators. The centralized estimator provides the best accuracy achievable with the

Figure 3: (a) Undirected communication topology $\mathcal{G}$ considered in the example scenario. (b) Illustration of the corresponding Markov Random Field $\mathcal{G}_X$ subject to estimation by the decentralized estimation network.

communication cost of collecting the network-wide measurements at a designated center. In the myopic estimation strategy, all variables are estimated locally using only the local measurements and no communication resources are utilized.

5.1. A Simple Gaussian Example

We first consider a small network composed of four platforms. A Gaussian random field $X = (X_1, X_2, X_3, X_4)$ is of concern and platform j is associated with X_j. We consider two-stage strategies over the undirected graph given in Figure 3(a). The BW constraints are captured by specifying the set of admissible symbols $\mathcal{U}_{i\to j} = \{0, 1, 2\}$ for all $(i, j) \in \mathcal{E}$.

The online processing, as described in Section 2.1, starts with each node evaluating its communication function on its measurement, i.e., nodes 1 − 4 simultaneously evaluate

$$u_{1\to3} = \mu_1(y_1),\ u_{2\to3} = \mu_2(y_2),\ (u_{3\to1}, u_{3\to2}, u_{3\to4}) = \mu_3(y_3),\ u_{4\to3} = \mu_4(y_4)$$

respectively. As soon as all the messages from the neighbors are received, estimation rules are run, i.e., nodes 1 − 4 evaluate

$$\hat{x}_1 = \nu_1(y_1, u_{3\to1}),\ \hat{x}_2 = \nu_2(y_2, u_{3\to2}),\ \hat{x}_3 = \nu_3(y_3, u_{1\to3}, u_{2\to3}, u_{4\to3}),\ \hat{x}_4 = \nu_4(y_4, u_{3\to4})$$

respectively. We design the strategy $\gamma = (\gamma_1, ..., \gamma_4)$ where $\gamma_j = (\mu_j, \nu_j)$ using Algorithm 3.

We select the communication cost local to node j as $c^c_j(u_{j\to ne(j)}, x_j) = \sum_{k\in ne(j)} c^c_{j\to k}(u_{j\to k}, x_j)$ which satisfies Assumption 1. Here, $c^c_{j\to k}(u_{j\to k})$ is the cost of transmitting the symbol $u_{j\to k}$ on the link $(j, k) \in \mathcal{E}$ and given by

$$c^c_{j\to k}(u_{j\to k}, x_j) = \begin{cases} 0, & \text{if } u_{j\to k} = 0 \\ 1, & \text{otherwise.} \end{cases}$$

Hence, $\mathcal{U}_{j\to k}$ together with $c^c_{j\to k}$ defines a selective communication scheme where $u_{j\to k} = 0$ indicates no communications and $u_{j\to k} \neq 0$ indicates transmission of a one bit message. We call this a 1-bit selective communication scheme and also discuss a 2-bit scheme later in this section. The estimation error is penalized by $c^d_j(x_j, \hat{x}_j) = (x_j - \hat{x}_j)^2$. Hence the total cost of a strategy is $J(\gamma) = J_d(\gamma) + \lambda J_c(\gamma)$ where J_d is the MSE and J_c is the total link use rate.

The random field prior is a multivariate Gaussian, i.e., $x \sim \mathcal{N}(x; \mathbf{0}, \mathbf{C}_X)$ where $\mathcal{N}$ denotes a multivariate Gaussian with mean $\mathbf{0}$ and covariance $\mathbf{C}_X$. This distribution is Markov with respect to the graph $\mathcal{G}_X$ in Figure 3(b). The covariance matrix is given by

$$\mathbf{C}_X = \begin{bmatrix} 2 & 1.125 & 1.5 & 1.125 \\ 1.125 & 2 & 1.5 & 1.125 \\ 1.5 & 1.5 & 2 & 1.5 \\ 1.125 & 1.125 & 1.5 & 2 \end{bmatrix}. \tag{35}$$

Note that Algorithm 3 is valid for any arbitrary selection of the undirected communication topology that is not necessarily identical to the Markov random field representation of X. Here, for the sake of simplicity we select the UG topology in Figure 3(a) to have the same structure as the MRF in Figure 3(b).

For the noise processes n_j for $j \in \mathcal{V}$, Assumptions 2 and 3 hold with $p(y_j|x_j) = \mathcal{N}(y_j; x_j, 0.5)$. Considering $\mathbf{C}_X$, each sensor has an SNR of 6dB.

The initial local estimation rule is the myopic minimum MSE estimator which is based only on y_j, i.e., $\nu^0_j(y_j, \overleftarrow{u}_j) = \int_{-\infty}^{\infty} \mathrm{d}x_j\, x_j p(x_j|y_j)$, and the initial communication rule is a threshold rule quantizing y_j given by

$$\mu^0_j(y_j) = \begin{cases} 1 \,, \, y_j < -2\sigma_n \\ 0 \,, \, -2\sigma_n \leqslant y_j \leqslant 2\sigma_n \\ 2 \,, \, y_j > 2\sigma_n \,. \end{cases} \tag{36}$$

Suppose that we use Algorithm 2 and achieve the performance points $(J_c(\gamma^*), J_d(\gamma^*))$ for the converged strategies as we vary λ. There exists a λ^* value such that for $\lambda \geq \lambda^*$, the communication cost λJ_c will increase to a level that prevents the decrease in the decision cost J_d achieved by the transmitted information among nodes to further cause a decrease in J. In this regime, not sending any messages (selecting the symbol 0) and using the myopic estimation rule will be the pbp optimal strategy. Hence, it is possible to interpret λ^* as the maximum price per bit that the system affords to decrease the expected estimation error. As we use Algorithm 3 and increase λ from 0 we approximate samples from the corresponding pareto-optimal curve which enables us to quantify the tradeoff between the cost of estimation errors and communication.

In Figure 4(a), we present the approximate MSE-total link use rate pairs of the converged strategies $\tilde{\gamma}^*$ obtained by using Algorithm 3 for varying λ from 0 with 0.001 steps (black '+'s). These points demonstrate graceful degradation of the estimation accuracy with decreasing communication load in the network. Specifically, we generate 2000 and 30000 samples from $p(x_i)$ and $p(y_i)$, respectively for obtaining S_{x_i} and S_{y_i}. The upper and lower bounds are MSEs corresponding to the myopic rule and the centralized optimal rule respectively. For the squared error cost, the optimal centralized rule given by $E\{X|Y = y\}$ yields a communication cost of $J_c = 3Q$ where Q is the number of bits used to represent a real number, i.e., y_j, before transmitting to the fusion center. Let us consider $(\tilde{J}_c, \tilde{J}_d)$ pairs for the 1-bit selective communication scheme, for $\lambda = 0$ (the transmission has no cost). The link use rate is approximately 3.2 bits, which is far less than the total capacity of 6 bits for the bi-directional topology given in Figure 3(a). Nevertheless, the MSE achieved by using the strategy designed using Algorithm 3 is significantly close to that for the centralized rule. The communication stops across the network for the strategy designed using $\lambda^* \approx 0.3$ and the nodes proceed with the myopic estimators for larger values of λ.

At this point, it is worth mentioning that the converged strategies for different threshold selections in the initial communication rule given by Eq.(36) yield the same performance with a slight variation due to Monte Carlo approximations. This indicates that the proposed scheme performs fairly consistently with different initializations, in this example.

We repeat the same scenario with a different BW constraint: Specifically, we select $\mathcal{U}_{i\to j}$s corresponding to a 2-bit selective communication scheme. The initial communication rules are appropriately modified versions of that given by Eq.(36) and the approximate performance points obtained are presented in Figure 4(a) as well[12]. The tradeoff curves show that, as we increase the link capacities and for small enough λ values, the pbp optimal strategies for the 2-bit case achieve fair improvements in the estimation accuracy for the same total communication load.

5.2. *A Simple Heavy Tailed Example*

In this example, we demonstrate that the MC framework applies for arbitrary distributions provided that samples can be generated from their marginals. This can be an important advantage in certain problem settings in which it is not possible to obtain closed form expressions even for the centralized rule. We

[12] For these experiments, we use the condition $\left\| \tilde{J}(\tilde{\gamma}^{l-1}) - \tilde{J}(\tilde{\gamma}^{l})\right| - \left|\tilde{J}(\tilde{\gamma}^{l-2}) - \tilde{J}(\tilde{\gamma}^{l-1})\right\| < 1.0e-2$ in the `Check` step of Alg. 3. The minimum number of iterations for convergence is 3 for both the 1- and 2-bit schemes and the resulting averages (standard deviations) are 3.24(0.43) and 3.11(0.31) for the 1- and 2-bit schemes, respectively.

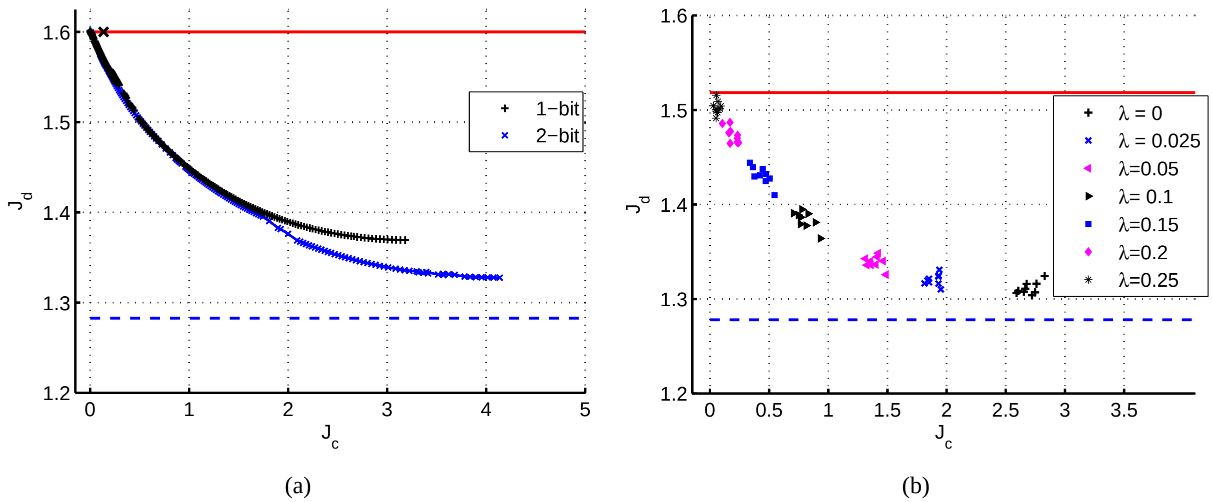


Figure 4: The approximate performance points converged revealing the tradeoff together with the lower bounds (blue dashed-lines) and the upper bounds (red dashed-lines) of the problems given by the estimation performance measured in MSE for the optimum centralized and the myopic rules respectively. (a) Gaussian UG problem: The estimation network in Figure 3(a) is subject to optimization through Algorithm 3. The initial strategy achieves $(J_c(\gamma^0), J_d(\gamma^0))$ (black 'x'). The pareto-optimal performance curves, achieved for the approximate pbp optimal strategies while λ is increased from 0 with steps of 0.001, are approximated by $\{(\tilde{J}_c(\tilde{\gamma}^*_\lambda), \tilde{J}_d(\tilde{\gamma}^*_\lambda))\}$ where $\tilde{\gamma}^*_\lambda$ is the approximated optimum strategy for λ. Results for 1 and 2 bit selective communication schemes are presented. (b) Heavy tailed (Laplacian) prior problem with a UG: We demonstrate the variation of the approximation over different sample sets for a heavy tailed prior through the performance points achieved using Alg. 3 with various values of λ and 10 sample sets for each λ.

consider such a scenario in which X is distributed by a heavy tailed prior $p(x)$, specifically a multivariate-symmetric Laplacian (MSL) given by

$$p(x) = \frac{2}{(2\pi)^{d/2}|C_x|^{1/2}} \left(\frac{x^T C_x^{-1} x}{2} \right)^{1-d/2} K_{1-d/2}(\sqrt{2x^T C_x^{-1} x}) \tag{37}$$

where d is the dimension of x, C_x is a covariance matrix, and $K_\eta(u)$ is the Bessel function of the second kind of order η (see, e.g., [39]). Let us denote this density by $SL_d(\mathbf{C}_X)$. Unlike the Gaussian case, uncorrelatedness does not imply independence and not being a member of the exponential family, $SL_d(\mathbf{C}_X)$ does not admit a Markov random field representation. On the other hand, it is possible to generate samples from an MSL utilizing samples generated from a multivariate Gaussian of zero mean and the desired covariance matrix together with samples drawn from the unit univariate exponential distribution, i.e., given $x' \sim \mathcal{N}(x'; \mathbf{0}, \mathbf{C}_X)$ and $z \sim e^{-z}$, generate samples of X by $x = \sqrt{z}x'$, then $x \sim SL_d(\mathbf{C_x})$.

Similar to that in the previous section, we assume the underlying communication structure described by

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in Figure 3(a) together with a 1-bit selective communication scheme, and similar cost functions, observation likelihoods, and initial local rules. To the best knowledge of the authors, for an MSL prior and Gaussian likelihoods, even the centralized paradigm fails to provide a solution without employing numerical approximations.

We consider $X = (X_1, X_2, X_3, X_4)$ such that $p_X(x) = SL_4(\mathbf{C}_X)$ where $\mathbf{C}_X$ is given by Eq.(35) and we exploit the fact that the j^{th} marginal density of $SL_d(\mathbf{C}_X)$ is given by $SL_1([\mathbf{C}_X]_{j,j})$. It is straightforward to generate samples from these marginals [40]. Sample sets from the observation distributions are obtained using the scheme in [20].

In this example, we also demonstrate the variation of the results over different sample sets, so, we generate 10 different sample sets such that $\left|S_j\right| = 3000$ and $\left|Q_j\right| = 45000$. Using these sets, we run Algorithm 3 for different choices of λ (as opposed to using a single sample set and small increments of λ as in Section 5.1). In Figure 4(b), approximate performance points for the converged strategies are presented. The upper and lower bounds are the MSEs corresponding to the myopic and the centralized rules, respectively[13]. For each value of λ, collective results based on the 10 sample sets provide a sample-based approximation to the performance point $(J_d(\gamma^*), J_c(\gamma^*))$ on the tradeoff curve[14]. These sample-based results form clusters with reasonable variability which can be interpreted as an indication of their approximation quality. It is reasonable to expect this level of variability since heavy tailed distributions require utilization of larger sample sets. Nevertheless, the proposed MC framework provides distributed solutions in problem settings which do not admit straightforward solutions even in the centralized case.

5.3. *Examples with Large Graphs*

In this section, we demonstrate Algorithm 3 in relatively large scale random field estimation problems. Specifically, we consider problems set up by randomly deploying 50 platforms over an area of 100 unit squares. Each sensor location $s_j \in \mathbb{R}^2$ is associated with a scalar random variable, X_j. We assume that the random field $X = (X_1, X_2, ..., X_{50})$ is Gaussian with zero mean, i.e., $x \sim \mathcal{N}(x; \mathbf{0}, \mathbf{C}_x)$ and $\mathbf{C}_x = [C_{i,j}]$ is

[13] In the MSL prior-Gaussian likelihoods problem, the evaluation of the myopic and centralized strategies and the corresponding MSEs require numerical approximations for which we utilize MC methods as well.

[14] Note that, $(J_d(\gamma^*), J_c(\gamma^*))$ is the performance of the pbp optimal strategy γ^* for the Bayesian risk corresponding to λ, i.e., $J(\gamma^*) = J_d(\gamma^*) + \lambda J_c(\gamma^*)$.

selected as the Matérn covariance with nugget effect given by [41]

$$C_{i,j} = \begin{cases} \left(\sigma^2/2^{(\eta-1)}\Gamma(\eta)\right)\left(2\sqrt{\eta}h/\phi\right)^{\eta} 2K_{\eta}\left(2\sqrt{\eta}h/\phi\right), h > 0 \\ \tau^2 + \sigma^2, h = 0 \end{cases} \tag{38}$$

where $h \triangleq \|s_i - s_j\|$ is the distance between sensors i and j, K_η is a modified Bessel function of the second kind of order η, τ^2 is the nugget effect, ϕ is the effective covariance range and σ^2 is referred to as the partial sill[15]. The covariance function for the particular set of parameter values we use in our experiments can be seen in Figure 5(b). The variances of X_js are given by the covariance function evaluated at $h = 0$ which is unity. The covariance matrix $\mathbf{C}_x$ for the deployment in Figure 5(a) is given in Figure 5(c). The inverse of $\mathbf{C}_x$ contains no zeros and, hence, this model cannot be exactly represented by a sparse Markov Random Field.

The undirected communication topology in Figure 5(a) is found by sparsifying the Gabriel graph of the deployment. We consider a one-bit selective bi-directional communication scheme which yields 128 bits total capacity with this UG. We initialize the nodes with *quantization* rules for communications and myopic estimators. We select a communication cost similar to that we have used in the previous examples and squared error as the estimation cost. We use $|S_j| = 2000$ and $|Q_j| = 30000$ samples from local marginals in Algorithm 3.

The measurement noise for each sensor is Gaussian with variance $\sigma^2_{n_j} = 0.25$ leading to $6.02dB$ signal-to-noise ratio (SNR) given by $\text{SNR} = 10\log_{10}\sigma^2_j/\sigma^2_{n_j}$. The myopic MSE is given by $\text{MSE} = \sigma^2_j\sigma^2_{n_j}/(\sigma^2_j + \sigma^2_{n_j})$ which equals to 0.2. In order to demonstrate the efficacy of the optimized two-stage strategies in comparison with the myopic estimator and the centralized estimator, we define an MSE equivalent SNR as $\text{SNR} = 10\log_{10}(\sigma^2_j - MSE)/MSE$. This quantity, in a sense, is the SNR of a sensor which would yield the given MSE value when it is used with a myopic estimator. From this viewpoint, a two-stage estimation strategy can be viewed as being equivalent to replacing each sensor with its SNR-improved version in a myopic strategy.

We consider sensors $17-23$ in Figure 5(a). In Figure5(d) we present the benefits of the two-stage strategies designed using Algorithm 3 in terms of the improvement in the MSE equivalent SNRs for different values of λ. The upper bounds are achieved by the centralized estimator. Nodes $18-23$ have closely located neighbors with highly correlated local variables. As λ is decreased, communication is utilized more, and, consequently an improvement as much as more than half of the myopic-centralized SNR gap is achieved. Node 17 is more distant to its neighbors and benefits less from the incoming information.

[15] Various forms of Matérn covariances are commonly used in spatial data modeling [8].

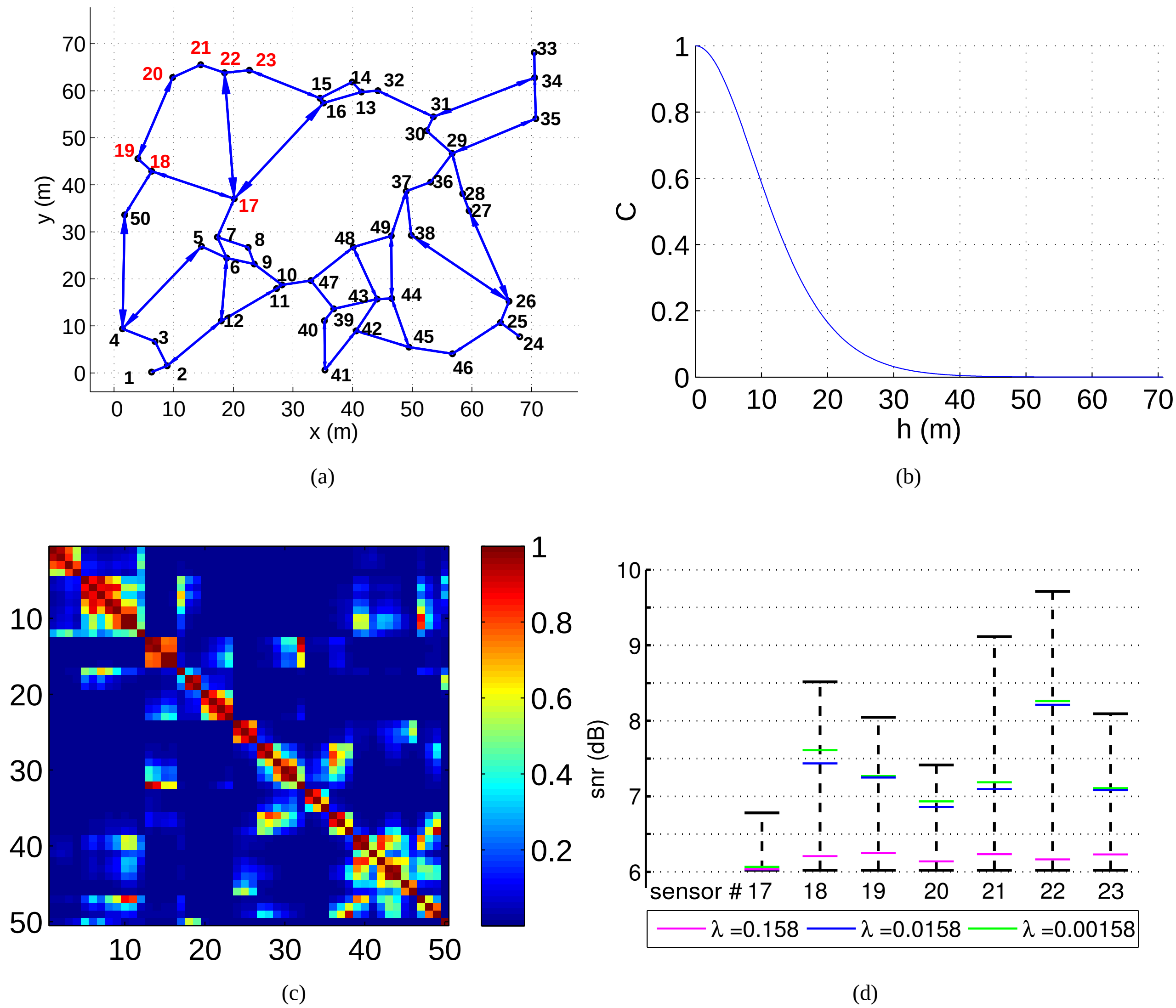


Figure 5: Set up for experiment involving 50 randomly deployed nodes: (a) Randomly distributed sensor nodes and the UG communication topology obtained by sparsifying the Gabriel Graph of the deployment. (b) Matérn covariance function (Eq.38) used in the experiments ($\tau^2 = \sigma^2 = 0.5, \nu = 4, \phi = 15$). (c) $\mathbf{C}_x$ obtained for the deployment in (a) with the covariance function in (b). (d) The myopic and centralized-equivalent SNRs of sensors $17-23$, and improvements achieved by optimizing the two-stage strategy with Algorithm 3 for different values of λ.

The overall estimation and communication costs of this network are given in Figure 6(a) for different values of λ and five different sample sets for each. Note that, the cost of communication for the improvement upon the myopic MSE is on the scale of tens of bits which is extremely small as compared to the cost of collecting network wide measurements at a designated node for centralized estimation. The performance points for different sample sets form clusters around the points from the pareto-optimal curve they approximate in a way similar to the example in Section 5.2 and the results given in Figure 4(b). The variations of the clusters indicate a fairly good quality of approximation. We verify the consistency of our algorithm in the performance of the designs by using four additional deployments. For each deployment, the diagonal of $\mathbf{C}_x$, and, hence, the myopic performances are the same with that of the other three networks. The MSEs of the centralized rules, on the other hand, differ as well as the total network capacities[16]. We present the approximate MSE-total link use rate points for $\lambda = 0.005$ and 0.05 and for 5 different sample sets in Figure 6(b)[17]. It can be observed that, the converged strategy improves the MSE performance in comparison with the myopic rule for all of the UGs with a fair amount of variability in the results. This suggests that our algorithm performs consistently across a variety of random network structures. The gains in the estimation accuracy in this example are fairly significant considering that only 1-bit transmissions are used. Our experiments also show that λ effectively controls the trade-off between estimation accuracy measured with MSE and the communications load in bits in large scale problems as well.

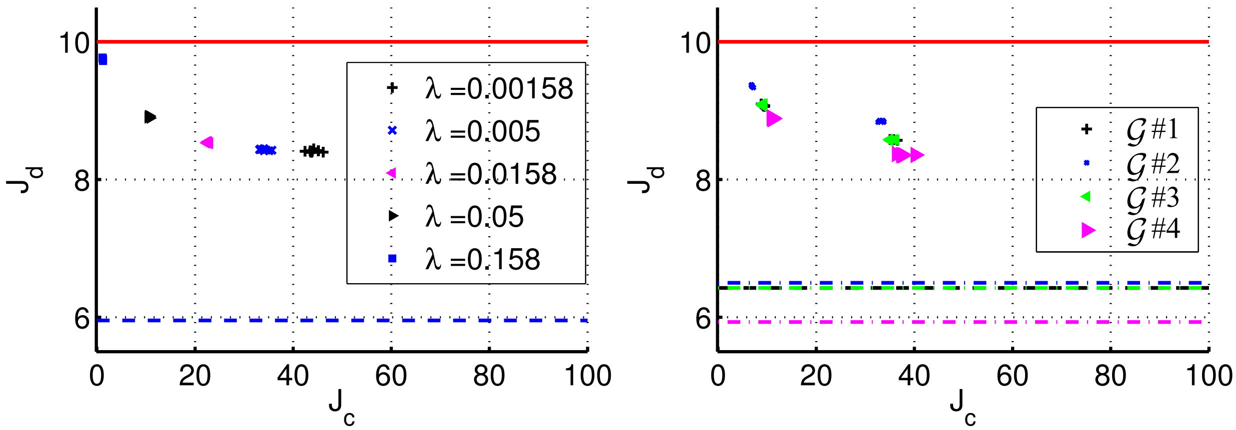
6. Conclusion

In this work, we have been concerned with the design of decentralized random field estimation strategies for sensor network applications. We constrain the feasible set of *online* strategies by the availability and BW of the links and use a design objective which allows us to trade the (possibly energy) price for communication off with the estimation accuracy. Person-by-person (pbp) optimal solutions to such problems can be found using *offline* iterative message passing algorithms which fit well into our context. In estimation problems, however, the optimization procedure as well as the pbp optimal local rules involve integral operators which cannot be evaluated exactly, in general. We have introduced a Monte Carlo framework which circumvents this problem and leads to a feasible decentralized optimization scheme while preserving the message passing structure. The proposed algorithm features scalability with the number of platforms as well as the number of variables involved. We have demonstrated these features through several examples including a Gaussian problem, a non-Gaussian prior problem, and random large graph scenarios. We have presented trade-off curves relating the MSE of estimation and the network wide communication load in bits.

One possible extension of this work is to investigate such strategies in settings involving broadcast communications with the nearest neighbors, unreliable channels, latency, sparse measurements and estimation

[16] The capacities corresponding to the deployment instances UG 1–4 are 132, 130, 134, and 140 bits, respectively.

[17] The number of iterations for convergence has a minimum value of 3, a mean value 4, and a standard deviation of 1.1.



(a) (b)

Figure 6: Algorithm 3 for five random UGs and for five sample sets for each deployment:(a) Performance points obtained for the UG in Figure 5(a). (b) Performance points obtained for four additional random UGs. The parameter λ is selected as $\lambda = 0.005, 0.05$ considering a 1-bit selective communication scheme and squared error estimation error penalty for all of the nodes. Note that the myopic MSE (showed by a solid red-line) is the same for all deployments whereas the centralized MSE (the lower bound) varies for each deployment.

of a random field over a grid. Another line of investigation would be to consider settings in which the random field prior evolves as a Markov process. Different in-network processing strategies can also be developed such as the hybrid in-network processing strategies (see [42] for such a perspective on the detection problem) employing both the class of strategies considered in this paper and strategies over DAGs [20]. It might also be worthwhile to consider the problem of selecting the communication graph structure that yields the best pbp optimal strategy given an *a priori* distribution.

Acknowledgment

The authors would like to thank Dr. O. Patrick Kreidl for his help and support during many discussions.

References

[1] D. Estrin, L. Girod, G. Pottie, M. Srivastava, Instrumenting the world with wireless sensor networks, in: Proceedings of the 2001 IEEE International Conference on Acoustic, Speech and Signal Processing, 2001.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, Computer Networks 38 (2002) 393–422.

[3] Smart sensor networks: Technologies and applications for green growth, OECD Report DSTI/ICCP/IE(2009)4/FINAL (December 2009).
URL http://www.oecd.org/dataoecd/39/62/44379113.pdf

[4] G. J. Pottie, W. Kaiser, Wireless integrated network sensors, Communications of the ACM 43 (5) (2000) 51–58.

[5] C.-Y. Chong, S. P. Kumar, Sensor networks: Evolution, opportunities, and challenges, Proceedings of the IEEE 91 (8) (2003) 1247–1256.

[6] W. R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in: Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000.

[7] P. S. Bernard, J. M. Wallace, Turbulent Flow: Analysis, Measurement and Prediction, John Wiley &Sons, 2002.

[8] H. Wackernagel, Multivariate Geostatistics, 3rd Edition, Springer, 2003.

[9] O. Schabenberger, C. A. Gotway, Statistical Methods for Spatial Data Analysis, Chapman & Hall/CRC Press, 2005.

[10] R. Nowak, U. Mitra, R. Willett, Estimating inhomogeneous fields using wireless sensor networks, Selected Areas in Communications, IEEE Journal on 22 (6) (2004) 999 – 1006. doi:10.1109/JSAC.2004.830893.

[11] Y. Wang, P. Ishwar, Distributed field estimation with randomly deployed, noisy, binary sensors, IEEE Transactions on Signal Processing 57 (3) (2009) 1177–1189.

[12] A. Dogandzic, K. Qiu, Decentralized random-field estimation for sensor networks using quantized spatially correlated data and fusion center feedback, IEEE Transactions on Signal Processing 56 (12) (2008) 6069–6085.

[13] H. Zhang, J. Moura, B. Krogh, Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost, IEEE Transactions on Signal Processing 57 (6) (2009) 2383–2395.

[14] J. Cortez, Distributed kriged kalman filter for spatial estimation, IEEE Transactions on Automatic Control 54 (12) (2009) 2816–2827.

[15] M. Çetin, L. Chen, J. W. Fisher III, A. T. Ihler, R. L. Moses, M. J. Wainwright, A. S. Willsky, Distributed fusion in sensor networks: A graphical models perspective, IEEE Signal Processing Magazine 23 (4) (2006) 42–55.

[16] A. T. Ihler, J. W. Fisher III, A. S. Willsky, Loopy belief propagation: Convergence and effects of message errors, Journal of Machine Learning Research 6 (2006) 905 – 936.

[17] M. Uney, M. Cetin, Graphical model-based approaches to target tracking in sensor networks: An overview of some recent work and challenges, in: Image and Signal Processing and Analysis, 2007. ISPA 2007. 5th International Symposium on, 2007, pp. 492–497. doi:10.1109/ISPA.2007.4383743.

[18] O. P. Kreidl, Graphical models and message-passing algorithms for network-constrained decision problems, Ph.D. thesis, MIT (2008).

[19] O. Kreidl, A. Willsky, An efficient message-passing algorithm for optimizing decentralized detection networks, IEEE Transactions on Automatic Control 55 (3) (2010) 563 –578. doi:10.1109/TAC.2009.2039547.

[20] M. Üney, M. Çetin, Monte Carlo optimization of decentralized estimation networks over directed acyclic graphs under communication constraints, IEEE Transactions on Signal Processing 59 (11) (2011) 5558 – 5576. doi:10.1109/TSP.2011.2163629.

[21] O. P. Kreidl, A. S. Willsky, Decentralized detection in undirected network topologies, in: Proceedings of the IEEE 14th Workshop on Statistical Signal Processing, SSP '07, 2007, p. 650=654.

[22] M. Üney, M. Çetin, An efficient Monte Carlo approach for optimizing decentralized estimation networks constrained by

undirected topologies, in: Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on, 2009, pp. 485 –488. doi:10.1109/SSP.2009.5278534.

[23] P. Santi, Topology control in wireless ad hoc sensor networks, ACM Computing Surveys 87 (2) (2005) 164–194.

[24] A. Bachir, M. Dohler, T. Watteyne, K.K. Leung, MAC Essentials for Wireless Sensor Networks, IEEE Communications Surveys & Tutorials, 12 (12) (2010) 222–248.

[25] B. Krishnamachari, Networking Wireless Sensors, Cambridge University Press, 2005, Ch. 8.

[26] J. Gao, L. Guibas, Geometric algorithms for sensor networks, Philosophical Transactions of the Royal Society A 370 (1958) (2012) 27–51.

[27] M. Yarvis, W. Ye, Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems, CRC Press, 2005, Ch. 13, Tiered Architectures in Sensor Networks.

[28] W. Su, E. Çayırcı, Özgür B. Akan, Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems, CRC Press, 2005, Ch. 16, Overview of Communication Protocols for Sensor Networks.

[29] J. N. Tsitsiklis, Advances in Statistical Signal Processing, JAI Press, 1993, Ch. Decentralized Detection, pp. 297–344.

[30] Y.-C. Ho, K.-C. Chu, Team decision theory and information structures in optimal control problems, IEEE Transactions on Automatic Control 29 (1) (1972) 22 – 28.

[31] S. P. Lloyd, Least Squares Quantization in PCM, IEEE Transactions on Information Theory IT-28 (2) (1982) 129–137.

[32] J. Max, Quantization for Minimum Distortion, IRE Transactions on Information Theory, 6 (1), (1960) 7–12.

[33] W.-M. Lam, A. R. Reibman, Design of quantizers for decentralized estimation systems, IEEE Transactions on communications 41 (11) (1993) 1602–1605.

[34] J. A. Gubner, Distributed estimation and quantization, IEEE Transactions on Information Theory 39 (4) (1993) 1456–1459.

[35] J. Geweke, Bayesian inference in econometric models using Monte Carlo integration, Econometrica 57 (6) (1989) 1317–1339.

[36] C. P. Robert, G. Casella, Monte Carlo Statistical Methods, 2nd Edition, Springer, 2004.

[37] E. B. Sudderth, A. T. Ihler, M. Isard, W. T. Freeman, A. S. Willsky, Nonparametric belief propagation, Commun. ACM 53 (10) (2010) 95–103. doi:10.1145/1831407.1831431.
URL http://doi.acm.org/10.1145/1831407.1831431

[38] A. T. Ihler, D. McAllester, Particle belief propagation, in: Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS), 2009, pp. 256–263.

[39] M. Abramowitz, I. A.Stegun (Eds.), Handbook of Mathematical Functions, Dover Publications, Inc., 1965, page 376.

[40] S. Kotz, T. J. Kozubowski, K. Podgorski, The Laplace Distribution and generalizations, Birkhauser, 2001.

[41] H.-R. Song, M. Fuentes, S. Ghosh, A comparative study of Gaussian geostatistical models and Gaussian Markow random field models, Journal of Multivariate Analysis 99 (8) (2008) 1681–1697.

[42] O. P. Kreidl, A. S. Willsky, Decentralized detection with long-distance communication, in: Proceedings of the 42nd Asilomar Conference on Signals, Systems and Computers, 2008, pp. 1352–1356.