Monte Carlo Optimization Approach for Decentralized Estimation Networks Under Communication Constraints

Murat Üney and Müjdat Çetin

Abstract

We consider designing decentralized estimation schemes over bandwidth limited communication links with a particular interest in the tradeoff between the estimation accuracy and the cost of communications due to, e.g., energy consumption. We take two classes of in–network processing strategies into account which yield graph representations through modeling the sensor platforms as the vertices and the communication links by edges as well as a tractable Bayesian risk that comprises the cost of communications and penalty for the estimation errors. This perspective captures a broad range of possibilities for "online" processing of observations as well as the constraints imposed and enables a rigorous design setting in the form of a constrained optimization problem. Similar schemes as well as the structures exhibited by the solutions to the design problem has been studied previously in the context of decentralized detection. Under reasonable assumptions, the optimization can be carried out in a message passing fashion. We adopt this framework for estimation, however, the corresponding optimization schemes involve integral operators that cannot be evaluated exactly in general. We develop an approximation framework using Monte Carlo methods and obtain particle representations and approximate computational schemes for both classes of in–network processing strategies and their optimization. The proposed Monte Carlo optimization procedures operate in a scalable and efficient fashion and, owing to the non-parametric nature, can produce results for any distributions provided that samples can be produced from the marginals. In addition, this approach exhibits graceful degradation of the estimation accuracy asymptotically as the communication becomes more costly, through a parameterized Bayesian risk.

Index Terms

Decentralized estimation, communication constrained inference, random fields, message passing algorithms, graphical models, Monte Carlo methods, wireless sensor networks, in-network processing, collaborative signal and information processing.

Murat Üney is with the Computer Vision and Pattern Analysis Lab., Signal and Information Processing Group, Sabancı University, Orhanlı-Tuzla 34956 İstanbul, Turkey (e-mail: muratuney@sabanciuniv.edu).

Müjdat Çetin is with the Faculty of Engineering and Natural Sciences, Sabancı University, Orhanlı-Tuzla 34956 İstanbul, Turkey (e-mail: mcetin@sabanciuniv.edu).

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 Young Scientist Award.

I. INTRODUCTION

The introduction of wireless sensor networks and their envisioned applications has nurtured the research on decentralized versions of canonical statistical inference problems in signal processing including detection, estimation and fusion. Typically, a large amount of observations induced by multiple quantities of interest are collected by sensor platforms at distinct locations and possibly in various modes [1]. While this spatially distributed nature neccessitates some communications, it is often the case that the components rely on limited energy stored in batteries [2], and transmitting bits is usually far more costly than computing them in terms of energy dissipation [3]. There are also resource limitations regarding sensing and computations and, therefore, any feasible processing scheme needs to take the relevant tradeoffs into account and ensure a collaborative operation of the components [4].

This work is motivated by the interest in designing decentralized processing schemes for estimation subject to a number of constraints regarding communications. The platforms setup a connected ad–hoc network on which it is possible to establish links between any two nodes and maintain higher level topologies yielding multi-tier architectures (see, e.g., [5]–[7]). These links are of finite capacity constraining the set of feasible symbols that can be transmitted over them and vary in length in the number of hops. The tradeoff between estimation accuracy and the cost of these transmissions is of concern to us. One possible way to abstract the energy cost of communications is to consider the number of hops and utilize a first order radio model for each hop, i.e., a model of energy dissipation for transmitting and receiving k bits at d meters distance (see e.g. [8]).

The phenomenon to be sensed is modeled by a collection of spatially correlated random variables. Such random-field models have been proposed in a variety of contexts including turbulent flow (Chp. 12 of [9]) and geostatistics data [10] such as temperature measurements over a field (Chp. 1 of [11]).

Previous work on decentralized estimation includes the canonical approach that assumes a star topology and bandwidth (BW) limited links in which a fusion center (FC) performs the estimation task based on messages from a finite alphabet sent by the so-called peripheral sensors. The transmitted symbols are quantized measurements and the design of quantizers together with a fusion rule is of concern in order to improve the estimation accuracy in various settings including Bayesian (e.g., [12], [13]), non-Bayesian (e.g., [14]), unknown prior and/or noise distribution (e.g., [15]–[17]), vector valued parameter (e.g., [18]) as well as the estimation of a random field (e.g., [19]–[21]). These treatments are limited in capturing certain aspects of the problem. First of all, the communication structures for which results can be produced are restricted to star topologies. Furthermore, the cost of transmissions from peripherals to the FC which possibly varies considering the multi–hop nature is not explicitly accounted for. Finally, often, a common random variable is of concern and estimation is performed only at the FC. This restricts the amount of collaboration among platforms for online processing of observations and opens up a possibility for a computational bottleneck in the case of multiple random variables (or a vector valued state) which can possibly be distributed over the nodes. We address these limitations through two classes of in–network processing strategies which capture a much broader range of communication and computation structures.

The decentralized random field estimation strategy in [19] utilizes bi-directional communications over a star topology and narrows the interval of uncertainty regarding the common variable based on reciprocal messaging between the FC and the peripherals. However, the variable representing the decision on the partition selection does not provide conditional independence for the observations, and exact fusion of messages is not tractable (which is carried out approximately using Monte Carlo approximations). Time-evolving random field estimation/prediction through Kalman-Bucy filtering (KBF) is considered in [22] and [23]. In particular, [23] addresses decentralized estimation through distributing the realization of the KBF, whereas [22] considers a center for filtering and communication constraints through surrogate communication costs and an estimation penalty. In order to reduce the amount of transmissions to the FC, model reduction is performed by variable selection at each step in a combinatorial setting. The problem we consider differs from this work in that, rather than considering a dynamical problem involving the processing of observations collected at consecutive time steps due to dynamical state transitions and modifying the model of the static estimation problem arising at each time step, we are interested in a static problem and optimization of a broader class of strategies such that graceful degradation is featured addressing the tradeoff.

Graphical models together with message passing algorithms has proved useful for decentralized statistical inference in sensor networks (see e.g., [24] and the references therein). In this framework, efficient statistical inference is achieved through message passing algorithms over a graph representation that reveals the probabilistic model underlying the estimation problem, which is often distinct from any graph representation of the available links. After mapping the former onto the latter, a decentralized inference scheme is obtained which can be realized provided that the underlying communication network supports the required messaging. It is often the case that the BW limitations necessitate approximations of the messages and consequently degrade the inference performance. Although it is possible to analyze the effects of these errors to some extent [25], it is hard to solve the problem of designing in–network processing schemes while taking into account the available links and capacities together with the cost of transmission over them (see, e.g., Chp. 5 of [26]).

We consider two classes of in–network processing strategies that are composed of local communication and computation rules and operate over a subset of all available communication links. For the first class, a directed acyclic graph (DAG) is rendered through the following: Treating the set of platforms as the vertex set of a graph, each node is associated with a (set of) random variable(s) from the collection, possibly with the variable(s) of a random field that model the phenomenon of interest at the location of the platform. Each link is represented by a directed edge starting from the source and terminating at the sink node. In addition, a set of admissible symbols that comply with the link capacity is associated with each edge. Given a set of links that renders a directed acyclic graph, a strategy is achieved by having all nodes produce outgoing messages to their children and an estimate of the random variable they are associated with, based on the incoming messages from their parents as well as the measurements they receive. Given a *prior* distribution for the random field and a tractable cost, this class yields a tractable Bayesian risk under a number of reasonable assumptions.

The second class allows bi–directional communications and considering edge pairs between two nodes that can perform peer–to–peer communications, renders an Undirected Graph (UG). Similar to that for the in–network

strategies over DAGs, each link is associated with a number of symbols according to the BW but, in contrast, local processing of nodes take place in two–stages. In the first stage, each node delivers messages to their neighbors based on its measurement. In the second stage, having received messages from their neighbors, each node perform estimation based on both the incoming messages and its measurement. One of the reasons for a two–stage strategy is to avoid possible deadlocks in the processing of the observations. Second, the assumptions that guarantee a tracktable Bayesian risk in the DAG case is not sufficient for strategies over UGs but the structure introduced by two–stage processing renders them sufficient.

As a result, both classes of strategies yield rigorous designs problems for decentralized inference under communication constraints in the form of constrained optimization problems in which the objective functions are Bayesian risks that penalize 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 links.

These classes of strategies together with the structures exhibited by the solutions have been recently studied in [27] (see also [28]–[31]) in the context of decentralized detection. For each class, after a Team Decision Theoretic investigation, an iterative procedure is obtained which, starting from an initial strategy, converges to a person–by–person optimal one and can be realized as a message passing algorithm, provided that certain assumptions hold.

We adopt this framework for decentralized *estimation* in which the variables of concern take values from denumerable sets, and hence yield expressions with integral operators that cannot be evaluated exactly in general. In order to keep the fidelity to the problem setting, we introduce an approximation framework utilizing Monte Carlo (MC) methods such that particle representations and approximate computational schemes for the operators replace the original expressions in both the strategies and their optimization. As a result, the iterative solutions are transformed to MC optimization algorithms which also maintain the following benefits of the original scheme: 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 and can be realized as message passing algorithms matching a possible self-organization requirement, provided that certain assumptions hold. Lastly, since the approach is Bayesian, it is possible to introduce information on the process of concern through a prior density function. In addition, the MC optimization schemes we propose feature 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 the marginals.

In the next section we introduce both classes of strategies, and then we define the problem in a constrained optimization setting. After presenting the Team Decision Theoretic investigation in Section III, we introduce our MC optimization framework for in–network processing strategies over DAGs and two–stage strategies over UGs in Sections IV and V respectively. Then we demonstrate the aforementioned features through several examples

in Section VI[1]. Finally we provide some observations together with possible future directions, and conclude in Section VII.

II. Problem Definition

We start this section with a number of basic definitions about our graphical representation of the problem and the variables involved in that representation. Then in Section II-A, we present the in–network processing paradigm over DAGs for "network constrained *online* processing" of the set of collected observations, which was previously studied in [27] for detection such that the elements of the earlier work (e.g., [34] [35] [35]) are unified including a DAG network topology, low–rate communication links between nodes and a spatially–distributed decision objective [31]. Then, in Section II-B, the two–stage strategies over UGs are introduced which enable modeling bi–directional links. Subsequently, in Section II-C, we state the design problem for the processing strategy taking into account communication constraints in a constrained optimization setting, which is to be solved *offline*, i.e., before processing the observations.

Common for both classes, a graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ represents an online communication and computation structure where each platform is associated with a node $v\in\mathcal{V}$. An edge $(i,j)\in\mathcal{E}$ corresponds to the finite capacity communication link from platform i to j on which i can transmit a symbol $u_{i\rightarrow j}$ without errors from the set of admissible symbols $\mathcal{U}_{i\rightarrow j}$. The number of elements in $\mathcal{U}_{i\rightarrow j}$, i.e., $|\mathcal{U}_{i\rightarrow j}|$, is finite and in accordance with the link capacity capturing the bandwidth constraints[2]. Note that, if $\mathcal{G}$ is a DAG, then a forward (backward) partial ordering is implied with respect to the reachability relation starting to count form the parentless (childless) nodes and proceeding forwards (backwards). If the links allow for bi–directional communication, i.e., $(i,j)\in\mathcal{E}$ implies that $(j,i)\in\mathcal{E}$, then $\mathcal{G}$ is an undirected graph.

We consider the joint probability distribution function (pdf) $P_{X,Y}(X,Y)$ where $X=(X_1,X_2,...,X_N)^T$ is the random variable to be estimated taking values from a denumerable set $\mathcal{X}=\mathcal{X}_1\times\mathcal{X}_2\times...\times\mathcal{X}_N$. Similarly $Y=(Y_1,Y_2,...,Y_M)^T$ takes values from a denumerable set $\mathcal{Y}=\mathcal{Y}_1\times\mathcal{Y}_2\times...\times\mathcal{Y}_M$ and is the collection of all observations induced by X. It holds that $N,M\geq 1$ and $\dim(\mathcal{X}_j),\dim(\mathcal{Y}_k)\geq 1$ for $j=1,...,N$ and $k=1,...,M$ respectively. A node $v\in\mathcal{V}$ collects $Y_v\subseteq\{Y_1,...,Y_M\}$ and can be associated with $X_v\subseteq\{X_1,...,X_N\}$ for which case it estimates X_v. This mapping which distributes the observed state over nodes is arbitrary, in principle, and enables decentralized inference with a broad range of possibilities. For simplicity, we assume that there are N platforms, with $M=N$ observations, and given $u,v\in\mathcal{V}$, X_u and X_v are mutually exclusive for $u\neq v$ throughout.

A. In–network processing strategies over DAGs

We first consider the class of strategies over DAGs for which the graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ modeling the communication and computation structure is directed and acyclic. Let $u_{\pi(j)}$ denote the incoming messages to node j from its parent

[1]The preliminary results of the proposed schemes appear in [32] and [33].

[2]For example, it is possible to represent a link with capacity $\log_2 d_{ij}$ bits with, e.g., selecting $\mathcal{U}_{i\rightarrow j}$ such that $|\mathcal{U}_{i\rightarrow j}|=d_{ij}+1$ where $0\in\mathcal{U}_{i\rightarrow j}$ indicates no transmission and enables a message cencoring or selective communication scheme. In [27], communication link errors are also considered which we do not take into account throughout.

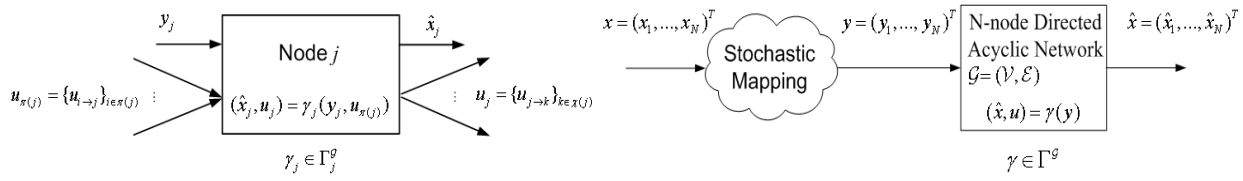


Fig. 1. Online processing scheme modeled with a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: (a) The viewpoint of node j in $\mathcal{G}$ which evaluates its local rule γ_j based on its measurement y_j as well as on the received messages $u_{\pi(j)}$ and produces an inference on the value of the random variable it is associated with, i.e., $\hat{x}_j$, together with outgoing messages u_j to its children. (b) The global view of the decentralized strategy over $\mathcal{G}$ where a random vector X takes the value x as the outcome of an experiment and induces observations y.

nodes $\pi(j)$, given by $u_{\pi(j)} \triangleq \{u_{i\to j} | i \in \pi(j)\}$. Let $\mathcal{U}_{\pi(j)}$ denote the set from which $u_{\pi(j)}$ takes values. This set is constructed through consecutive Cartesian products given by $\mathcal{U}_{\pi(j)} \triangleq \underset{i\in\pi(j)}{\otimes} \mathcal{U}_{i\to j}$ where $\otimes$ denotes consecutive Cartesian Products[3]. The set of outgoing messages from node j to child nodes $\chi(j)$, given by $u_j \triangleq \{u_{j\to k} | k \in \chi(j)\}$, takes values from the set $\mathcal{U}_j$ which can be defined in a similar way to that for $\mathcal{U}_{\pi(j)}$ as $\mathcal{U}_j \triangleq \underset{k\in\chi(j)}{\otimes} \mathcal{U}_{j\to k}$.

As node j measures $y_j \in \mathcal{Y}_j$ and receives $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$; it evaluates a function, called its local rule, defined by

$$\gamma_j : \mathcal{Y}_j \times \mathcal{U}_{\pi(j)} \to \mathcal{U}_j \times \mathcal{X}_j$$

which produces an estimate $\hat{x}_j \in \mathcal{X}_j$ as well as outgoing messages $u_j \in \mathcal{U}_j$. The design process of the optimal γ_j is the topic of Section II-C. The space of rules local to node j is given by

$$\Gamma_j^{\mathcal{G}} \triangleq \{\gamma_j | \gamma_j : \mathcal{Y}_j \times \mathcal{U}_{\pi(j)} \to \mathcal{U}_j \times \mathcal{X}_j\}$$

where the superscript $\mathcal{G}$ denotes that the definition of the set relies on $\mathcal{G}$. Considering the space of all possible estimators, i.e., $\Gamma \triangleq \{\gamma | \gamma : \mathcal{Y} \to \mathcal{X}\}$, it holds that $\Gamma^{\mathcal{G}} \subset \Gamma$. Note that $\{\mathcal{U}_{i\to j} | (i,j) \in \mathcal{E}\}$ also relies on $\mathcal{G}$ through the edge set $\mathcal{E}$.

A DAG implies a partial ordering and it is possible to obtain a forward and backward partial ordering in accordance with the reachability relation such that the parentless and the childless nodes have the smallest order respectively. The directed acyclic nature of $\mathcal{G}$ leads to causal online processing of the observations when the nodes execute their local rules in accordance with the forward partial order, i.e., starting from the parentless nodes, at each step, nodes with the corresponding order evaluate their local rules and processing stops after the childless nodes. The process from node j's point of view is illustrated in Fig. 1(a).

Considering $\mathcal{V} = \{1, 2, ..., N\}$, the aggregation of local rules denoted by γ is called a strategy, i.e., $\gamma = (\gamma_1, \gamma_2, ..., \gamma_N)$, and takes values from the set of feasible strategies given by

$$\Gamma^{\mathcal{G}} = \Gamma_1^{\mathcal{G}} \times \Gamma_2^{\mathcal{G}} \times ... \times \Gamma_N^{\mathcal{G}}$$

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

which will simply be denoted by $\Gamma^{\mathcal{G}} = \underset{v\in\mathcal{V}}{\otimes} \Gamma_v^{\mathcal{G}}$. The set of all messages in the network arising for the "online" processing of the observations is given by $u \triangleq \{u_{i\to j}|(i,j)\in\mathcal{E}\}$, and takes values from $\mathcal{U} \triangleq \underset{(i,j)\in\mathcal{E}}{\otimes} \mathcal{U}_{i\to j}$. The global view of this paradigm is illustrated in Fig. 1(b).

B. Two–stage in–network processing strategies over UGs

Given a UG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, it holds for all edges in $\mathcal{G}$, i.e., $(i,j)\in\mathcal{E}$, that $(i,j)\in\mathcal{E} \Leftrightarrow (j,i)\in\mathcal{E}$ establishing a bi–directional setting. Unlike the DAG case, the local rules operate in two–stages: In the first stage, having observed $y_j \in \mathcal{Y}_j$, node j transmits a message $u_{j\to i}$ taking values from $\mathcal{U}_{j\to i}$ to each of its neighbors $i \in ne(j)$ constituting $u_j = \{u_{j\to i}|i\in ne(j)\}$. The set of all possible outgoing messages is given by $\mathcal{U}_j = \underset{i\in ne(j)}{\otimes} \mathcal{U}_{j\to i}$. In the second stage, an inference on the value of X_j is drawn based on the observation y_j and the incoming messages from neighboring nodes given by $u_{ne(j)} = \{u_{i\to j}|i\in ne(j)\}$. The set of all possible incoming messages is given by $\mathcal{U}_{ne(j)} = \underset{i\in ne(j)}{\otimes} \mathcal{U}_{i\to j}$.

A causal online processing of measurements takes place when each $j\in\mathcal{V}$, first performs its local communication rule $\mu_j : \mathcal{Y}_j \to \mathcal{U}_j$ acting on only y_j, and after $u_{ne(j)}$ is received, proceeds with the local decision rule $\nu_j : \mathcal{Y}_j \times \mathcal{U}_{ne(j)} \to \mathcal{X}_j$. Hence, the local rule of node j is a pair given by $\gamma_j = (\mu_j, \nu_j)$ and the design process of the optimal γ_j is the topic of Section II-C.

Similar to the discussion in the DAG case, it is possible to define the space of all first–stage (communication) rules as

$$\mathcal{M}_j^{\mathcal{G}} = \{\mu_j|\mu_j : \mathcal{Y}_j \to \mathcal{U}_j\}$$

and the second-stage (estimation) rule space by

$$\mathcal{N}_j^{\mathcal{G}} = \{\nu_j|\nu_j : \mathcal{Y}_j \times \mathcal{U}_{ne(j)} \to \mathcal{X}\}$$

The local rule spaces $\Gamma_j^{\mathcal{G}} = \mathcal{M}_j^{\mathcal{G}} \times \mathcal{N}_j^{\mathcal{G}}$ for $j\in\mathcal{V}$ construct the strategy space $\Gamma^{\mathcal{G}} = \underset{v\in V}{\otimes} \Gamma_v^{\mathcal{G}}$.

C. Design problem in a constrained optimization setting

For any such in–network processing strategy, it is possible to 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 u are assigned, i.e., $c : \mathcal{U}\times\mathcal{X}\times\mathcal{X}\to\mathbb{R}$. In addition, given $\gamma = (\gamma_1, ..., \gamma_N) \in \Gamma^{\mathcal{G}}$, the tuple $(U^T, \hat{X}^T)^T = \gamma(Y)$ is a random variable conditionally independent of X given Y, denoted by $(U^T, \hat{X}^T)^T \perp\!\!\!\perp X \,|\, Y$, and the distribution $p(u,\hat{x}|y)$ is specified by γ and denoted by $p(u,\hat{x}|y;\gamma)$. Note that, by construction, considering the causal online processing in the DAG and UG cases

$$p(u,\hat{x}|y;\gamma) = \prod_{j=1}^{N} p(u_j,\hat{x}_j|y_j,u_{\pi(j)};\gamma_j) \tag{1}$$

and

$$\begin{aligned} p(u,\hat{x}|y;\gamma) &= \prod_{j\in\mathcal{V}} p(u_j,\hat{x}_j|y_j,u_{ne(j)};\gamma_j) \\ &= \prod_{j\in\mathcal{V}} p(u_j|y_j;\mu_j)p(\hat{x}_j|y_j,u_{ne(j)};\nu_j) \end{aligned} \tag{2}$$

hold respectively.

Consider a Bayesian risk, i.e., $E\{c(u,x,\hat{x});\gamma\}$. The distribution used to perform the expectation operation is specified by γ and can be constructed through Eq. (1) and Eq. (2) for the strategies over DAGs and two–stage strategies over UGs respectively as

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

Therefore, for any given strategy $\gamma\in\Gamma^{\mathcal{G}}$, there corresponds a Bayesian risk and the problem of finding the best strategy for estimation under communication constraints described by $\mathcal{G}$ turns into a constrained optimization problem given by

$$\begin{aligned}\text{(P)}: \quad & \min J(\gamma) \\ & \text{subject to } \gamma\in\Gamma^{\mathcal{G}}\end{aligned} \tag{4}$$

where $J(\gamma)=E\{c(u,x,\hat{x});\gamma\}$.

It can be shown that if there exists an optimal strategy, then there exist an optimal deterministic strategy [36]. Therefore it suffices to consider the deterministic local rule spaces which consequently implies a treatment of the distribution $p(u_j,\hat{x}_j|y_j,u_{\pi(j)};\gamma_j)$ as a finite set of distributions parameterized on u_j in the DAG case, i.e.,

$$p(u_j,\hat{x}_j|y_j,u_{\pi(j)};\gamma_j)=p_{u_j}(\hat{x}_j|y_j,u_{\pi(j)};\gamma_j) \tag{5}$$

$$p_{\left[\gamma_j(y_j,u_{\pi(j)})\right]_{\mathcal{U}_j}}(\hat{x}_j|y_j,u_{\pi(j)};\gamma_j)=\delta(\hat{x}_j-\left[\gamma_j(y_j,u_{\pi(j)})\right]_{\mathcal{X}_j}) \tag{6}$$

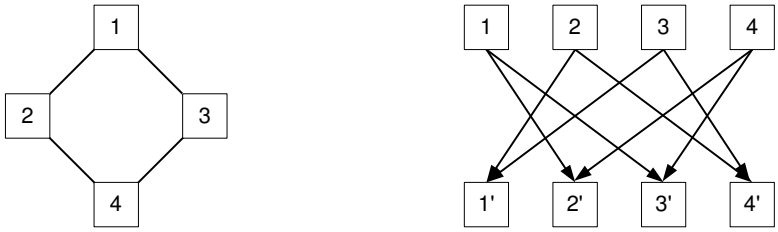
where we denote with $[.]_{\mathcal{X}}$ the element of its n-tuple argument that takes values from the set $\mathcal{X}$ and δ is Dirac's delta distribution. Hence, the local rule γ_j and the distribution family $p_{u_j}(\hat{x}_j|y_j,u_{\pi(j)};\gamma_j)$ specify each other accordingly. Moreover, Eq.(5) substituted in Eq.(1) constructs the distribution given by Eq.(3) which underlies Problem (P).

Similarly, for the two–stage strategies over UGs, the local first and second stage rules determine the following distributions

$$p(u_j|y_j;\mu_j)=\delta_{u_j,\mu_j(y_j)} \tag{7}$$

$$p(\hat{x}_j|y_j,u_{ne(j)};\nu_j)=\delta(\hat{x}_j-\nu_j(y_j,u_{\pi(j)})) \tag{8}$$

where $\delta_{i,j}$ is the Kronecker's delta. For the case, the distribution given by Eq.(3) is constructed by substituting Eq.s (7) and (8) in Eq.(2). It is also possible to express the two–stage in–network processing strategies by unwrapping the UG to a directed graph which is bipartite and hence acyclic [30]. Consider, for example, the undirected graph and its unwrapped directed counterpart in Fig. 2. Nodes $1-4$ perform only the stage-one communication rules, i.e., μ_js, and nodes $1'-4'$ are associated only with the stage-two estimation rules, i.e., ν_js. Node j and j' correspond to the same physical platform but different processing tasks, in this respect. The unwrapped counterparts enable us to apply the solutions to the design problem for the DAG case for two–stage strategies over UGs as well.



(a) (b)

Fig. 2. (a) A loopy UG of 4 nodes (b) the DAG counterpart regarding the two–stage online processing: Nodes 1–4 correspond to platforms 1–4 but only performing the first–stage communication rules, whereas nodes $1'$–$4'$ correspond to platforms 1–4 but only performing the second–stage estimation rules.

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

1: Choose $\gamma^0 = (\gamma_1^0, \gamma_2^0, ..., \gamma_N^0)$ such that $\gamma_j \in \Gamma_j^{\mathcal{G}}$ for $j = 1, 2, ..., N$; Choose $\varepsilon \in \mathbb{R}^+$;$l = 0$ ▷ `Initialize`

2: $l = l + 1$

3: `For` $j = N, N-1, \ldots, 1$ `Do` $\gamma_j^l = \arg\min_{\gamma_j \in \Gamma_j^{\mathcal{G}}} J(\gamma_1^{l-1}, ..., \gamma_{j-1}^{l-1}, \gamma_j, \gamma_{j+1}^l, ..., \gamma_N^l)$ ▷ `Update`

4: `If` $J(\gamma^{l-1}) - J(\gamma^l) < \varepsilon$ `STOP, else, GO TO` 2; ▷ `Check`

Note that, it is possible to express the treatment in [12], [13] as well as the bounded parameters estimation setting utilized in [14], [17] through a non–informative prior and a cost function c penalyzing only estimation errors, i.e., $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, within the framework above.

III. Team Decision Theoretic Investigation

Problem (P) in (4) is a typical team decision problem [37] and such problems are intractable in various settings, including conventional decentralized detection in which star–topologies are considered and $\mathcal{X}$ is finite [36]. Nevertheless, necessary (but not sufficient) conditions of optimality yield nonlinear Gauss-Seidel iterations which converge to a person–by–person optimal strategy. Given an optimal strategy $\gamma^* \in \Gamma^{\mathcal{G}}$ it holds that $J(\gamma_j^*, \gamma_{\setminus j}^*) \leq J(\gamma_j, \gamma_{\setminus j}^*)$ for all $\gamma_j \in \Gamma_j^{\mathcal{G}}$ where $\setminus j$ denotes $\mathcal{V} \setminus j$ and $\gamma_{\setminus j}^* = \{\gamma_1^*, \gamma_2^*, ..., \gamma_{j-1}^*, \gamma_{j+1}^*, ..., \gamma_N^*\}$ [4]. Equivalently a relaxation of (P) is to find a Nash equilibirium where no change in a single local rule yields a better objective value, i.e., one is interested in finding $\gamma^* = (\gamma_1^*, ..., \gamma_n^*)$ such that

$$\gamma_j^* = arg \min_{\gamma_j \in \Gamma_j} J(\gamma_j, \gamma_{\setminus j}^*) \tag{9}$$

for all $j \in \{1, 2, ..., N\}$. Such a solution is also said to be person–by–person (pbp) optimal and it is possible to converge to one starting from an initial strategy by the immediate iterations given by Algorithm 1.

Considering problem (P) in the detection setting, the optimal strategies from the classes of concern lie in a finitely parameterized subspace of $\Gamma^{\mathcal{G}}$ under certain conditions [28], [30] and consequently tractable "offline" optimization

[4]Note that, when it is obvious from the context, we abuse the notation and denote $\{x_i | i \in I\}$ by x_I where I is an index set for the collection of variables $\{x_1, x_2, ..., x_N\}$.

algorithms are obtained for both strategies over DAGs and two–stage strategies over UGs which operate in an iterative fashion. We adopt the elaborate investigation of Kreidl (Chp.s 3 and 4 in [27]) for decentralized estimation under communication constraints and obtain variational forms for the pbp optimal local rules which differ from that in the detection setting in that, functions over denumerable domains parameterize the pbp optimal local rules.

A. Pbp optimal in–network strategies over DAGs

In this Section, we present the pbp optimal strategies for in–network strategies over DAGs which are estimation counterparts of those in the detection setting together with conditions under which an efficient online processing is achieved [31].

The pbp optimal strategies exhibit certain structures provided certain assumptions hold. The first condition that leads a useful form for the pbp optimal local rules is the conditional independence of observations:

Assumption 1: (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., $p(x,y)=p(x)\prod_{i=1}^{N}p(y_i|x)$.

Proposition 3.1: (Proposition 3.1 in [27] for estimation) Consider (P) under Assumption 1. The $j^{\,th}$ pbp optimal rule given by Eq.(9) reduces to

$$\gamma_j^*(y_j,u_{\pi(j)})=\arg\min_{(u_j,\hat{x}_j)\in\mathcal{U}_j\times\mathcal{X}_j}\int_{\mathcal{X}}\mathrm{d}x\,p(y_j|x)\theta_j^*(u_j,\hat{x}_j,x;u_{\pi(j)}) \tag{10}$$

where

$$\theta_j^*(u_j,\hat{x}_j,x;u_{\pi(j)})=p(x)\sum_{u_{\setminus\{j\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus j}}\mathrm{d}\hat{x}_{\setminus j}\,c(u,\hat{x},x)\prod_{i\neq j}\int_{\mathcal{Y}_i}\mathrm{d}y_i\,p(y_i|x)p(u_i,\hat{x}_i|y_i,u_{\pi(i)};\gamma_i^*) \tag{11}$$

for all $u_{\pi(j)}\in\mathcal{U}_{\pi(j)}$ and $y_j\in\mathcal{Y}_j$ with non-zero probability, i.e., $p(y_j,u_{\pi(j)};\gamma_{\setminus j}^*)>0$.

Proof: The proof follows the factorization of $J(\gamma)=J(\gamma_j,\gamma_{\setminus j})$ after substituting $\gamma_{\setminus j}=\gamma_{\setminus j}^*$, Eq.s(1),(5),(6) and Assumption 1 together with the fact that if a pbp local rule exists, then a deterministic pbp local rule exists [36].

After substituting $\gamma_{\setminus j}=\gamma_{\setminus j}^*$, Eq.(1) and Assumption 1 in $J(\gamma)=J(\gamma_j,\gamma_{\setminus j})$ we obtain

$$\begin{aligned}
J(\gamma_j,\gamma_{\setminus j}^*) &= \int_{\mathcal{X}}\mathrm{d}x\int_{\mathcal{X}}\mathrm{d}\hat{x}\sum_{u\in\mathcal{U}}c(u,x,\hat{x})p(x)p(u_j,\hat{x}_j|x,u_{\pi(j)};\gamma_j)\prod_{i\neq j}^{N}p(u_i,\hat{x}_i|x,u_{\pi(i)};\gamma_i^*)\\
&= \int_{\mathcal{Y}_j}\mathrm{d}y_j\int_{\mathcal{X}_j}\mathrm{d}\hat{x}_j\sum_{u_j\in\mathcal{U}_j}\sum_{u_{\pi(j)}\in\mathcal{U}_{\pi(j)}}p(u_j,\hat{x}_j|y_j,u_{\pi(j)};\gamma_j)\int_{\mathcal{X}}\mathrm{d}x p(y_j|x)p(x)\\
&\qquad\sum_{u_{\setminus\{j\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus j}}c(u,x,\hat{x})\prod_{i\neq j}^{N}p(u_i,\hat{x}_i|x,u_{\pi(i)};\gamma_i^*)
\end{aligned} \tag{12}$$

Consider deterministic local rules such that $\gamma_j\in\Gamma_j^{\mathcal{G}}$ and Eq.s(5) and (6). Given $(u_{\pi(j)},y_j)\in\mathcal{U}_{\pi(j)}\times\mathcal{Y}_j$ with non-zero probability, γ_j^* minimizes Eq.(12) with probability 1 provided that for $(u_j^*,\hat{x}_j^*)$

$$p_{u_j}(\hat{x}_j|y_j,u_{\pi(j)};\gamma_j)=\begin{cases}\delta(\hat{x}_j-\hat{x}_j^*) & \text{, if } u_j=u_j^*\\ 0 & \text{, otherwise}\end{cases} \tag{13}$$

where the weight of $(u_j^*, \hat{x}_j^*)$ in Eq.(12), i.e.,

$$\int_{\mathcal{X}} \mathrm{d}x p(y_j|x)p(x) \sum_{u_{\setminus\{j\}\cup\pi(j)}} \int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j} c(u, x, \hat{x}_{\setminus j}, \hat{x}_j = \hat{x}_j^*) \prod_{i\neq j, i\notin\chi(j)} \int_{\mathcal{Y}_i} \mathrm{d}y_i p(u_i, \hat{x}_i|u_{\pi(i)}, y_i; \gamma_i^*)$$
$$\prod_{i\neq j, i\in\chi(j)} \int_{\mathcal{Y}_i} \mathrm{d}y_i p(u_i, \hat{x}_i|u_{j\to i}^* \cup \{u_{i'\to i|i'\in\pi(i)\setminus j}\}, y_i; \gamma_i^*) p(y_i|x) \quad (14)$$

is minimum. Hence, for all $(u_{\pi(j)}, y_j) \in \mathcal{U}_{\pi(j)} \times \mathcal{Y}_j$ with non-zero probability

$$\gamma_j^*(y_j, u_{\pi(j)}) = arg \min_{(u_j, \hat{x}_j)\in\mathcal{U}_j\times\mathcal{X}_j} \int_{\mathcal{X}} \mathrm{d}x p(y_j|x)\theta_j^*(u_j, \hat{x}_j, x; u_{\pi(j)})$$

where θ_j^* is identified as

$$\theta_j^*(u_j, \hat{x}_j, x, u_{\pi(j)}) = p(x) \sum_{u_{\setminus j}\in\mathcal{U}_{\setminus j}} \int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j}\, c(u, \hat{x}, x) \prod_{i\neq j} \int_{\mathcal{Y}_i} \mathrm{d}y_i\, p(y_i|x) p(u_i, \hat{x}_i|y_i, u_{\pi(i)}; \gamma_i^*) \quad (15)$$

■

Regarding Proposition 3.1 (and Eq.(10) in particular), it can be shown that

$$\int_{\mathcal{X}} \mathrm{d}x p(Y_j|x)\theta_j^*(u_j, \hat{x}_j, x; U_{\pi(j)}) \propto E\{c(u, x, \hat{x})|Y_j, U_{\pi(j)}; \gamma_{\setminus j}^*\}$$

where u_j and $\hat{x}_j$ are free variables[5] and in this respect it is revealed that the j^{th} pbp optimal rule involves minimizing the conditional expected cost given the incoming messages $u_{\pi(j)}$ and the measurement y_j where the underlying distribution is specified by all the local rules other than the j^{th}.

Note that in Eq.(10), θ_j^* does not depend on the observation y_j and the likelihood $p(y_j|x_j)$ acts as a sufficient statistics. Hence, θ_j provides a useful parameterization for the j^{th} pbp optimal rule, which, unlike its appearance as a finite dimensional vector in the detection setting [29], is a function over a denumerable domain. In addition, it is useful to treat the right hand side (RHS) of Eq. (11) as an operator ψ such that given any set of local rules for nodes other than the j^{th}, i.e., $\gamma_{\setminus j} \in \Gamma_{\setminus j}^{\mathcal{G}}$, fixed not necessarily at an optimum, ψ produces θ_j, i.e., $\theta_j = \psi_j(\gamma_{\setminus j})$. Then, the corresponding local rule for the j^{th} node is obtained through Eq.(10) which can also be treated as an operator given θ_j, i.e., $\gamma_j = \varsigma_j(\theta_j)$. Therefore, it is possible to obtain an iterative scheme which, starting from an initial strategy, converges to a pbp optimal one, in principle, by replacing the `Update` step of Algorithm 1 with

$$\theta_j^l = f_j(\theta_1^{l-1}, ..., \theta_{j-1}^{l-1}, \theta_{j+1}^l, ..., \theta_N^l) \quad (16)$$

for $j = 1, 2, ..., N$ where f_j denotes the composite operator (obtained after substituting $\varsigma_i(\theta_i)$ for all $i \in \setminus j$ in ψ_j). Note that, as a consequence of the fact that $\mathcal{X}$ is denumarable, the fixed point equations $\{\theta_j = f_j(\theta_{\setminus j})\}_{j\in\mathcal{V}}$ corresponding to Algorithm 1 with the aforementioned modification are not practically solvable in general.

Nevertheless, optimality in a pbp sense has been considered in the decentralized estimation literature for the canonical star–topology. For example, Proposition 3.1 applied for quantizer peripherals and a fusion center setting together with a squared error cost, i.e., $c(u, \hat{x}, x) = (\hat{x} - x)^2$, specializes to the optimality conditions presented in [12]. For this case, the structure of the local rules as given above do not yield closed form representations in general,

[5]Note that $c(u, x, \hat{x})$ can be expanded as $c((u_{\setminus j}, u_j), x, (\hat{x}_{\setminus j}, \hat{x}_j))$ to explicitly show the free variables u_j and $\hat{x}_j$ of the j^{th} local rule.

altough relatively straightforward numerical computations are involved when the joint density $p(x, y_1, ..., y_N)$ is Gaussian and x is a scalar. The fact that the fusion rule is not scalable in the number of peripherals raises the potential issue of computational bottlenecks. This consideration has led to a fusion rule which is linear in the received symbols [13].

1) Efficient online strategies: We continue with assumptions under which efficient online processing becomes possible [31]:

Assumption 2: (Measurement Locality) Every node j observes y_j due to only x_j, i.e., $p(y_j|x) = p(y_j|x_j)$.

Corollary 3.2: (Corollary 3.2 in [27] for Estimation) Under Assumptions 1 and 2, the j^{th} pbp optimal rule given by Proposition 3.1 reduces to

$$\gamma_j^*(Y_j, U_{\pi(j)}) = \arg \min_{u_j \times \hat{x}_j \in (\mathcal{U}_j \times \mathcal{X}_j)} \int_{\mathcal{X}_j} \mathrm{d}x_j p(Y_j|x_j) \phi^*(u_j, \hat{x}_j, x_j; U_{\pi(j)}) \tag{17}$$

where

$$\phi_j^*(u_j, \hat{x}_j, x_j; u_{\pi(j)}) = \int_{x_{\setminus j} \in \mathcal{X}_{\setminus j}} \mathrm{d}x_j \theta_j^*(u_j, \hat{x}_j, x; u_{\pi(j)}) \tag{18}$$

Proof: Substitute $p(y_j|x) = p(y_j|x_j)$ in Eq.(10) and rearrange the terms. ■

Under Assumptions 1 and 2, the local rules evaluate marginalizations over only the set from which the associated variable takes values from, i.e., $\mathcal{X}_j$, rather than $\mathcal{X}$, and become independent of the number of nodes. This provides scalability in the number of nodes (and correspondingly the number of variables) and hence efficiency for online processing.

2) Efficient offline optimization: Corollary 3.2 provides an efficient oline processing strategy. However, we do not have such efficiency for specifying the pbp optimal local rules since ϕ_j^* given by Eq.(18) depends on all the nodes other than the j^{th}. Under additional assumptions discussed below, the offline optimization scales with the number of nodes:

Assumption 3: (Cost Locality) The Bayesian cost function is additive over the nodes $j \in \mathcal{V}$, i.e.,

$$c(u, \hat{x}, x) = \sum_{j \in \mathcal{V}} c_j(u_j, \hat{x}_j, x_j) \tag{19}$$

Assumption 4: (Polytree Topology) Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a polytree, i.e., $\mathcal{G}$ is a directed acyclic graph with an acyclic undirected counterpart[6].

Proposition 3.3: (Proposition 3.2 in [27] for estimation) Consider Problem (P) given in (4) such that X and $\hat{X}$ take values from a denumerable set $\mathcal{X}$. Under Assumptions 1–4 , Eq.(17) applies with

$$\phi_j^*(u_j, \hat{x}_j, x_j; u_{\pi(j)}) \propto p(x_j) P_j^*(u_{\pi(j)}|x_j) \left[c_j(u_j, \hat{x}_j, x_j) + C_j^*(u_j, x_j) \right] \tag{20}$$

[6]Note that a polytree implies a forward (backward) partial–order starting from the parentless (childless) nodes with respect to the reachability relation.

where $P_j^*(u_{\pi(j)}|x_j)$ is the incoming message likelihood given by the forward recursion

$$P_j^*(u_{\pi(j)}|x_j) = \begin{cases} 1 & \text{, if } \pi(j) = \emptyset \\ \int\limits_{\mathcal{X}_{\pi(j)}} dx_{\pi(j)} p(x_{\pi(j)}|x_j) \prod\limits_{i \in \pi(j)} P_{i \to j}^*(u_{i \to j}|x_i) & \text{, otherwise} \end{cases} \tag{21}$$

with terms regarding influence of $i \in \pi(j)$ on j given by

$$P_{i \to j}^*(u_{i \to j}|x_i) = \sum_{u_{\chi(i)\setminus j} \in \mathcal{U}_{\chi(i)\setminus j}} \sum_{u_{\pi(i)} \in \mathcal{U}_{\pi(i)}} P_i^*(u_{\pi(i)}|x_i) \int\limits_{\mathcal{X}_i} \mathrm{d}\hat{x}_i \int\limits_{\mathcal{Y}_i} \mathrm{d}y_i \, p(u_i, \hat{x}_i | y_i, u_{\pi(i)}; \gamma_i^*) p(y_i|x_i) \tag{22}$$

and the conditional cost term $C_j^*(u_j, x_j)$ which is added to the local cost and given by the backward recursion

$$C_j^*(u_j, x_j) = \begin{cases} 0 & \text{, if } \chi(j) = \emptyset \\ \sum_{k \in \chi(j)} C_{k \to j}^*(u_{j \to k}, x_j) & \text{, otherwise} \end{cases} \tag{23}$$

with terms regarding the influence of $k \in \chi(j)$ on j given by

$$C_{k \to j}^*(u_{j \to k}, x_j) = \int\limits_{\mathcal{X}_{\pi(k)\setminus j}} dx_{\pi(k)\setminus j} \int\limits_{\mathcal{X}_k} dx_k p(x_{\pi(k)\setminus j}, x_k | x_j) \sum_{u_{\pi(k)\setminus j} \in \mathcal{U}_{\pi(k)\setminus j}} \prod_{m \in \pi(k)\setminus j} P_{m \to k}^*(u_{m \to k}|x_m) \times$$

$$I_k^*(u_{\pi(k)}, x_k; \gamma_k^*) \tag{24}$$

and

$$I_k^*(u_{\pi(k)}, x_k; \gamma_k^*) = \int\limits_{\mathcal{Y}_k} \mathrm{d}y_k \int\limits_{\mathcal{X}_k} \mathrm{d}\hat{x}_k \sum_{u_k \in \mathcal{U}_k} \left[c_k(u_k, \hat{x}_k, x_k) + C_k^*(u_k, x_k) \right] p(u_k, \hat{x}_k | y_k, u_{\pi(k)}; \gamma_k^*) p(y_k|x_k) \tag{25}$$

Proof: (Sketch) First, we recognize that the DAG structure together with Assumption 2 implies that the set of incoming messages $u_{\pi(j)}$ depends on not all the rules other than the j^{th} but only those of the ancestors of j denoted by $an(j)$, i.e., $p(u_{\pi(j)}|x; \gamma_{\setminus j}^*) = p(u_{\pi(j)}|x_{an}(j); \gamma_{an(j)}^*)$. Under Assumption 3 the output of the j^{th} local rule, i.e., $(u_j, \hat{x}_j)$, does not affect the costs of nodes other than the descendants of j denoted by $de(j)$, i.e.,

$$E\{\sum_{i \in \setminus j} c(u_i, \hat{x}_i, x_i) | u_j, \hat{x}_j; \gamma_{\setminus j}^*\} = E\{\sum_{i \in \setminus j \setminus de(j)} c(u_i, \hat{x}_i, x_i); \gamma_{\setminus j}^*\} + E\{\sum_{i \in de(j)} c(u_i, \hat{x}_i, x_i) | u_j, \hat{x}_j; \gamma_{\setminus j}^*\}$$

In other words, optimization of γ_j can be performed equivalently with an objective regarding the costs only on node j and its descendants. Under Assumption 4, the operation of rules local to the ancestors of j and descendants of j are mutually exclusive and the incoming message likelihoods and the expected costs yield the structure given by Eq.(20). Moreover, Assumption 4 guarantees that there are no parent nodes with common ancestors and no child nodes with common descendants yielding the multiplicative structure in Eq.s(21)–(22) and the additive structure of the expected costs in Eq.s(23)–(25). A detailed proof is provided in Appendix A. ■

Considering Eq.s(21) and (22) we note that $P_{i \to j}^*(u_{i \to j}|x_i)$ is the likelihood of x_i based on the particular message $u_{i \to j}$ on the link from node i to j and under Assumption 4 $P_j^*(u_{\pi(j)}|x_j)$ is the likelihood of x_j for the particular incoming message vector $u_{\pi(j)}$, i.e., $p(u_{\pi(j)}|x_j; \gamma_{an(j)})$. A similar treatment of Eq.s(23) and (24) reveals that $C_{k \to j}^*(u_{j \to k}, x_j)$ terms are the expected cost if the actual value of the random variable associated with node j takes the value x_j and node j sends the message $u_{j \to k}$ on the link to its child k. Hence, under a polytree topology

$C_j^*(u_j, x_j)$ is the total expected cost induced on the descendats of j for transmitting u_j. This cost is added to the local cost $c_j(u_j, \hat{x}_j, x_j)$ in Eq.(20) which also penalizes the transmission cost. Also considering Eq.s(17) and (20), and noting that under these assumptions $p(x_j)p(y_j|x_j)P(u_{\pi(j)}|x_j) \propto p(x_j|y_j, u_{\pi(j)})$, we conclude that given the measurement y_j and the incoming messages $u_{\pi(j)}$, node j chooses the output with the minimum expected cost where this cost is the sum of the costs due to the local rule of node j and rules of its descendants and the underlying distribution is determined by the rules local to ascendants of node j.

Similar to the treatment regarding Proposition 3.1 to yield the set of fixed point equations given by Eq.17, it is possible to consider Eq.s (21)–(25) as operators for any given (not neccessarily optimal) strategy $\gamma_{\setminus j} \in \Gamma_{\setminus j}^{\mathcal{G}}$. Similarly, it is possible to summarize this treatment by d_j, f_j, g_j and h_j such that

$$\phi_j = d_j(P_j, C_{\chi(j)\to j}) \tag{26}$$

$$P_j = f_j(P_{\pi(j)\to j}) \tag{27}$$

$$P_{j\to\chi(j)} = g_j(\phi_j, P_j) \tag{28}$$

$$C_{j\to\pi(j)} = h_j(\phi_j, P_{\pi(j)\to j}, C_{\chi(j)\to j}) \tag{29}$$

where $P_{\pi(j)\to j} = \{P_{i\to j}\}_{i\in\pi(j)}$, $C_{\chi(j)\to j} = \{C_{k\to j}\}_{k\in\chi(j)}$ and $C_{j\to\pi(j)} = \{C_{j\to i}\}_{i\in\pi(j)}$. Note that d_j, f_j, g_j and h_j are specified by the RHSs of Eq.s(20) and (23), Eq.(21), Eq.(22), and finally Eq.s(24) and (25) respectively. Consequently, the forward recursion implied by f_j and g_j with respect to the forward partial–ordering of $\mathcal{G}$ together with the backward recursion implied by h_j and d_j with respect to the backward partial–ordering yields Algorithm 2 after replacing the `Update` step of Algorithm 1 as described.

It is possible to perform this algorithm in a message passing fashion treating each node $j \in \mathcal{V}$ as an entity which can perform computations and communications. Each node $j \in \mathcal{V}$ starts only with the knowledge of $p(x_j, x_{\pi(j)})$ and $c(u_j, \hat{x}_j, x_j)$ and an initial local rule $\gamma_j^0 \in \Gamma_j^{\mathcal{G}}$ which determines $p(u_j, \hat{x}_j|y_j, u_{\pi(j)}; \gamma_j^0)$. In the forward pass, starting from the parentless nodes and proceeding in forward partial ordering implied by $\mathcal{G}$, each node receives $P_{i\to j}$ from its parents $i \in \pi(j)$, computes $P_{j\to k}$ for its children $k \in \chi(j)$ and transmits them. In the backward pass, starting from the childless nodes and proceeding in the backward partial–ordering, each node receives $C_{k\to j}$ from its children $k \in \chi(j)$ and computes $C_{j\to i}$ for its parents $i \in \pi(j)$ which involves updating the local rule. Note that, in contrast with the online processing strategy which assumes a polytree topology allowing only uni–directional links, the message passing interpretation of the offline strategy optimization requires bi–directional communications. It is reasonable to assume that both the topology assumed by the online processing and the links required by the offline optimization are provided by the underlying network layer through physically available connections and appropriate protocols [5]–[7].

In Section III-A1, owing to the information structure introduced under Assumptions 1 and 2, an efficient online processing strategy is achieved. With the addition of Assumptions 3–4, the optimization of the local rules in a pbp sense admits a message passing algorithm which scales both with the number of variables and the number of platforms. The resulting iterative scheme given as Algorithm 2 is amenable for network self-organization and for a

Algorithm 2 Iterations converging to a pbp optimal in-network processing strategy over a DAG $\mathcal{G}$.

1: Choose $\gamma^0 = (\gamma_1^0, \gamma_2^0, ..., \gamma_N^0)$ such that $\gamma_j^0 \in \Gamma_j^{\mathcal{G}}$ for $j = 1, 2, ..., N$; Choose $\varepsilon \in \mathbb{R}^+$;$l = 0$ ▷ Initialize

2: $l = l + 1$

3: For $j = 1, 2, ..., N$ Do ▷ Update Step 1: Forward Pass

$$P_j^l = f_j(\{P_{i\to j}^l(u_{i\to j}|x_i)\}_{i\in\pi(j)})$$

$$\{P_{j\to k}^l(u_{j\to k}|x_j)\}_{k\in\chi(j)} = g_j(\phi_j^{l-1}, P_j^l)$$

4: For $j = N, N-1, ..., 1$ Do ▷ Update Step 2: Backward Pass

$$\phi_j^l = d_j(P_j^l, \{C_{k\to j}^l(u_{j\to k}, x_j)\}_{k\in\chi(j)})$$

$$\{C_{j\to i}^l(u_{i\to j}, x_i)\}_{i\in\pi(j)} = h_j(\phi_j^l, \{P_{i\to j}^l(u_{i\to j}|x_i)\}_{i\in\pi(j)}, \{C_{k\to j}^l(u_{j\to k}, x_j)\}_{k\in\chi(j)})$$

5: If $J(\gamma^{l-1}) - J(\gamma^l) < \varepsilon$ STOP, else GO TO 2 ▷ Check

network that would execute the resulting strategy for a certain amount of time after initialization, the communication cost of the optimization procedure might be considered as reasonable [31].

It is often the case that it is hard to achieve consistency in penalizing the estimation errors and communication costs through an arbitrary selection of the cost function $c : \mathcal{U} \times \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. It is possible to select one which results in smooth degradation in the estimation performance as the link utilization is decreased. Also considering Proposition 3.3, we assume a separable cost and develop the simplifications this provides.

Assumption 5: (Separable Costs) The global cost function $c(u, \hat{x}, x)$ is separable to functions penalizing estimation errors and communications. In particular, $c(u, \hat{x}, x) = c^d(\hat{x}, x) + \lambda c^c(u, x)$ where c^d and c^c are cost functions for estimation errors and communications respectively. Here, λ appears as a unit conversion constant and can be interpreted as the equivalent estimation penalty per unit communication cost [27]. 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[7].

Note that, Assumption 5, together with Assumption 3 implies that the local cost functions are separable, i.e.,

$$c_j(u_j, x_j, \hat{x}_j) = c_j^d(x_j, \hat{x}_j) + \lambda c_j^c(u_j, x_j) \tag{30}$$

Corollary 3.4: Consider Proposition 3.3, if the local costs are separable, i.e., Assumption 5 holds in addition to Assumptions 1-4, then the pbp optimal local rule in the variational form given by Eq.(17) is separated into two

[7]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 with λ.

rules for estimation and communication as $\gamma_j^* = (\nu_j^*, \mu_j^*)$ given by

$$\hat{x}_j = \nu_j^*(y_j, u_{\pi(j)}) = \arg\min_{\hat{x}_j \in \mathcal{X}_j} \int_{x_j \in \mathcal{X}_j} dx_j p(x_j) p(y_j|x_j) P_j^*(u_{\pi(j)}|x_j) c_j^d(\hat{x}_j, x_j) \tag{31}$$

$$u_j = \mu_j^*(y_j, u_{\pi(j)}) = \arg\min_{u_j \in \mathcal{U}_j} \int_{x_j \in \mathcal{X}_j} dx_j p(x_j) p(y_j|x_j) P_j^*(u_{\pi(j)}|x_j) \left[\lambda c_j^c(x_j, u_j) + C_j^*(u_j, x_j)\right] \tag{32}$$

Moreover, the corresponding distribution $p(u_j, \hat{x}_j|y_j, u_{\pi(j)}; \gamma_j^*)$ given by Eq.(5) takes the form

$$p(u_j, \hat{x}_j|y_j, u_{\pi(j)}; \gamma_j^*) = p(\hat{x}_j|y_j, u_{\pi(j)}; \nu_j^*) p(u_j|y_j, u_{\pi(j)}; \mu_j^*) \tag{33}$$

Proof: After substituting the separable local cost in Eq.(20) and Eq.(17), the optimization is separated into two problems over arguments $\hat{x}_j \in \mathcal{X}$ and $u_j \in \mathcal{U}_j$. This separation also implies that U_j and $\hat{X}_j$ are conditionally idependent denoted by $U_j \perp\!\!\!\perp \hat{X}_j \,|\, (Y_j, U_{\pi(j)})$ yielding Eq.(33) by definition. ■

Example 3.5: Consider a separable local cost where the estimation penalty is given by $c_j^d(\hat{x}_j, x_j) = (\hat{x}_j - x_j)^2$ as in the conventional mean squared error (MSE) estimator. We obtain a closed form expression for the estimation rule regarding the variational form in Eq.(31) after differentiating with respect to $\hat{x}$ and setting the result equal to zero:

$$\hat{x}_j = \nu_j^*(Y_j, U_{\pi(j)}) = \frac{\int_{\mathcal{X}_j} \mathrm{d}x_j \, x_j p(x_j) p(Y_j|x_j) P_j^*(U_{\pi(j)}|x_j)}{\int_{\mathcal{X}_j} \mathrm{d}x_j \, p(x_j) p(Y_j|x_j) P_j^*(U_{\pi(j)}|x_j)} \tag{34}$$

Note that the information structure implies that $P_j^*(u_{\pi(j)}|x_j) = p(u_{\pi(j)}|x_j; \gamma_{\setminus j}^*)$ holds which in turn is equal to $p(u_{\pi(j)}|x_j; \gamma_{an(j)}^*)$ due to the polytree topology. In addition, the conditional independence relation $U_{\pi(j)} \perp\!\!\!\perp Y_j \,|\, X_j$ holds such that equivalently $p(x_j, y_j, u_{\pi(j)}) = p(x_j) p(y_j|x_j) p(u_{\pi(j)}|x_j)$. Hence the denominator in Eq.(34) is nothing but $p(y_j, u_{\pi(j)}) = p(y_j, u_{\pi(j)}; \gamma_{an(j)}^*)$ and the estimator is given by

$$\hat{x}_j = \nu_j^*(y_j, u_{\pi(j)}) = \int_{\mathcal{X}_j} \mathrm{d}x_j \, x_j p(x_j|y_j, u_{\pi(j)}; \gamma_{an(j)}^*)$$

which is the center of gravity of the posterior density conditioned on both the observation and the incoming messages (this density is specified by the rules local to ancestors of j, i.e., $\gamma_{an(j)}^*$, under Assumptions 1-4, which are fixed at the optimum). Hence, any selection of the communication rules for ancestors manifest themselves in the optimal estimation rule for node j through the likelihood $P_j^*(u_{\pi(j)}|x_j)$. Under this particular choice of the decision cost, $u_{\pi(j)}$ is treated as another conditionally independent observation while utilizing the MSE estimator based on the posterior.

If the local cost functions are separable, similar simplifications to those in Proposition 3.3 take place.

Corollary 3.6: Consider Proposition 3.3, if the local costs are separable, then $I^*(u_{\pi(k)}, x_k; \gamma_k^*)$ given by Eq.(25) takes the form

$$I^*(u_{\pi(k)}, x_k; \gamma_k^*) = J_{d|x_k, u_{\pi(k)}} + J_{c|x_k, u_{\pi(k)}} \tag{35}$$

where $J_{d|x_k, u_{\pi(k)}}$ is the local expected estimation cost conditioned on x_k and $u_{\pi(k)}$ given by

$$J_{d|x_k, u_{\pi(k)}} = \int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k \, c_k^d(\hat{x}_k, x_k) p(\hat{x}_k|x_k, u_{\pi(k)}; \nu_k^*) \tag{36}$$

and $J_{c|x_k,u_{\pi(k)}}$ is the total expected cost of transmitting the symbol u_k conditioned on x_k and $u_{\pi(k)}$, including costs induced on the descendats, i.e., $C_k^*(u_k,x_k)$, as well as the transmission cost captured by $c_k^c(u_k,x_k)$, i.e.,

$$J_{c|x_k,u_{\pi(k)}} = \sum_{u_k\in\mathcal{U}_k} \left(\lambda c_k^c(u_k,x_k) + C_k^*(u_k,x_k)\right) p(u_k|x_k,u_{\pi(k)};\mu_k^*) \tag{37}$$

Moreover, the conditional pdf of the estimations specified by ν_k^* is given by

$$p(\hat{x}_k|x_k,u_{\pi(k)};\nu_k^*) = \int_{\mathcal{Y}_k} \mathrm{d}y_k\, p(\hat{x}_k|y_k,u_{\pi(k)};\nu_k^*)p(y_k|x_k) \tag{38}$$

and the conditional pmf of the outgoing messages specified by μ_k^* is given by

$$p(u_k|x_k,u_{\pi(k)};\mu_k^*) = \int_{\mathcal{Y}_k} \mathrm{d}y_k\, p(u_k|y_k,u_{\pi(k)};\mu_k^*)p(y_k|x_k) \tag{39}$$

Proof: After substituting the separable local cost for node k given by Eq.(30) in Eq.(25) and rearranging terms

$$\begin{aligned} I_k^*(u_{\pi(k)},x_k;\gamma_k^*) = \int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k c_k^d(\hat{x}_k,x_k) \int_{\mathcal{Y}_k} \mathrm{d}y_k p(\hat{x}_k|y_k,u_{\pi(k)};\nu_k^*)p(y_k|x_k) \\ + \lambda \sum_{u_k\in\mathcal{U}_k} \left[\lambda c_k^c(u_k,x_k) + C_k^*(u_k,x_k)\right] \int_{\mathcal{Y}_k} \mathrm{d}y_k p(u_k|y_k,u_{\pi(k)};\mu_k^*)p(y_k|x_k) \end{aligned} \tag{40}$$

is obtained. ■

Therefore, under Assumptions 1– 5, sufficient conditions of optimality in a pbp sense are provided by Eq.s (20)–(24) together with Eq.s (35)–(39) implying an iterative optimization scheme. In principle, once the operators implied by these expressions are utilized in Algorithm 2, it is possible to find a pbp optimal decentralized estimation strategy starting with an initial one.

Finally, the corresponding Bayesian risk at the l^{th} step, i.e., $J(\gamma^l)$, which is also required by the `Check` step of Algorithm 2 is obtained as

$$J(\gamma^l) = \sum_{j\in\mathcal{V}} G_j(\gamma_j^l) \tag{41}$$

where

$$\begin{aligned} G_j(\gamma_j^l) = \int_{\mathcal{X}_j} \mathrm{d}x_j p(x_j) \sum_{u_{\pi j}\in\mathcal{U}_{\pi j}} P_j^{l+1}(u_{\pi(j)}|x_j) \int_{\mathcal{Y}_j} \mathrm{d}y_j \int_{\mathcal{X}_j} \mathrm{d}\hat{x}_j \sum_{u_j\in\mathcal{U}_j} c_j(u_j,\hat{x}_j,x_j) \\ p(u_j,\hat{x}_j|y_j,u_{\pi(j)};\gamma_j^l)p(y_j|x_j) \end{aligned} \tag{42}$$

B. Pbp optimal two–stage in–network processing strategies over UGs

The information structure of the directed case yield the conditions given by Proposition 3.1 provided that Assumption 1 holds which specializes to Proposition 3.3 if additionally Assumptions 2-4 are satisfied. On the other hand, considering decentralized strategies constrained by an undirected graph, Proposition 3.1 applies to the unwrapped directed counterpart under Assumption 1 and the following [30]:

Assumption 6: The global cost function is the sum of costs due to the stage-one communication rules and stage-two decision rules, which are in turn additive over the nodes, i.e.,

$$c(u,\hat{x},x)=\sum_{i=1}^{N}\left[c_i^d(\hat{x}_i,x)+\lambda c_i^c(u_i,x)\right] \tag{43}$$

Note that, simultaneous satisfaction of Assumptions 3 and 5 is equivalent to simultaneous satisfaction of Assumptions 3 and 6. If Assumptions 1 and 5 hold together with Assumptions 2 and 3, then Proposition 3.3 applies to the unwrapped directed counterpart of the two–stage strategy over a UG [27] and the following holds:

Proposition 3.7: (Proposition 4.3 in [27] for estimation) Under Assumptions 1–3 and 5, $J(\gamma)=J_d(\gamma)+\lambda J_c(\gamma)$ holds and given a pbp optimal strategy $\gamma^*=(\gamma_1^*,...\gamma_N^*)$ constituted of two–stage local rules over an undirected graph and fixing all local rules other than the j^{th}, the j^{th} optimal rule reduces to local stage–one communication rule given by

$$\mu_j^*(y_j)=\arg\min_{u_j\in\mathcal{U}_j}\int_{\mathcal{X}_j}\mathrm{d}x_j p(y_j|x_j)\alpha_j^*(u_j,x_j) \tag{44}$$

where

$$\alpha_j^*(u_j,x_j)\propto p(x_j)[\lambda c_j^c(u_j,x_j)+C_j^*(u_j,x_j)] \tag{45}$$

for all $y_j\in\mathcal{Y}_j$ with nonzero probability and stage two–estimation rule given by

$$\nu_j^*(y_j,u_{ne(j)})=\arg\min_{\hat{x}_j\in\mathcal{X}_j}\int_{\mathcal{X}_j}\mathrm{d}x_j p(Y_j|x_j)\beta_j^*(x_j,\hat{x}_j,u_{ne(j)}) \tag{46}$$

where

$$\beta_j^*(x_j,\hat{x}_j,u_{ne(j)})\propto p(x_j)P_j^*(u_{ne(j)}|x_j)c_j^d(\hat{x}_j,x_j) \tag{47}$$

for all $y_j\in\mathcal{Y}_j$ and for all $u_{ne(j)}\in\mathcal{U}_{ne(j)}$ with nonzero probability.

The incoming message likelihood is given by

$$P_j^*(u_{ne(j)}|x_j)=\int_{\mathcal{X}_{ne(j)}}\mathrm{d}x_{ne(j)}p(x_{ne(j)}|x_j)\prod_{i\in ne(j)}P_{i\to j}^*(u_{i\to j}|x_i) \tag{48}$$

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

$$P_{i\to j}^*(u_{i\to j}|x_i)=\sum_{u_i\setminus u_{i\to j}}p(u_i|x_i;\mu_i^*) \tag{49}$$

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

$$p(u_i|x_i;\mu_i^*)=\int_{\mathcal{Y}_i}\mathrm{d}y_i p(y_i|x_i)p(u_i|y_i;\mu_i^*) \tag{50}$$

In addition for all $u_j\in\mathcal{U}_j$

$$C_j^*(u_j,x_j)=\sum_{i\in ne(j)}C_{i\to j}^*(u_{j\to i},x_j) \tag{51}$$

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

$$C_{i\to j}^*(u_{j\to i},x_j)=\int_{\mathcal{X}_{ne(i)\setminus j}}dx_{ne(i)\setminus j}\int_{\mathcal{X}_i}dx_i p(x_{ne(i)\setminus j},x_i|x_j)$$
$$\sum_{u_{ne(i)\setminus j}}\prod_{j'\in ne(i)\setminus j}P_{j'\to i}^*(u_{j'\to i}|x_{j'})I_i^*(u_{ne(i)},x_i;\gamma_i^*) \tag{52}$$

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

1: Choose $\gamma^0 = (\gamma_1^0, \gamma_2^0, ..., \gamma_N^0)$ such that $\gamma_j^0 \in \Gamma_j^{\mathcal{G}}$ for $j = 1, 2, ..., N$; Choose $\varepsilon \in \mathbb{R}^+$;$l = 0$ ▷ Initialize

2: $l = l + 1$

3: For $j = 1, 2, ..., N$ Do ▷ Update Step 1: Compute message likelihoods

$$P_{j\to ne(j)}^l = g_j(\alpha_j^{l-1})$$

4: For $j = 1, 2, ..., N$ Do ▷ Update Step 2: Update the stage--two rules

$$P_j^l = f_j(P_{ne(j)\to j}^l)$$

$$\beta_j^l = q_j(P_j^l)$$

$$C_{j\to ne(j)}^l = h_j(\beta_j, P_{ne(j)\to j}^l)$$

5: For $j = 1, 2, ..., N$ Do ▷ Update Step 3: Update the stage--one rules.

$$\alpha_j^l = r_j^l(C_{ne(j)\to j}^l)$$

6: If $J(\gamma^{l-1}) - J(\gamma^l) < \varepsilon$ STOP, else GO TO 2 ▷ Check

such that

$$I_i^*(u_{ne(i)}, x_i; \nu_i^*) = \int_{\mathcal{Y}_i} \mathrm{d}y_i \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i \, c_i^d(\hat{x}_i, x_i) p(\hat{x}_i|y_i, u_{ne(i)}; \nu_i^*) p(y_i|x_i) \tag{53}$$

Proof: Apply Corollary 3.3 on the unwrapped directed couterpart of the undirected graph $\mathcal{G}$ together with the two–stage local rules. Note that the j^{th} pbp optimal local rule given in Proposition 3.3 reduces to the form given in Corollary 3.4 under Assumption 5 which is implied by Assumptions 3 and 6. ■

Through Proposition 3.7, given a person-by-person optimal strategy, we obtain stage–one communication and stage–two estimation rules local to node j in a variational form, based on the rules local to the the remaining nodes. Considering Eq.s(48) and (49), $P_j^*(u_{ne(j)}|x_j)$ is the likelihood of x_j given $u_{ne(i)}$. Eq.s(51)-(53) reveal that $C_j^*(u_j, x_j)$ is the total expected cost induced on the neighbors by u_j, i.e., $E\{c^d(\hat{x}_{ne(j)}, x_{ne(j)})|u_j, x_j; \gamma_{\setminus j}^*\}$. Since $p(x_j)p(y_j|x_j)P(u_{ne(j)}|x_j) \propto p(x_j|y_j, u_{ne(j)})$ holds under Assumptions 1-3 and 5, 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 minimum expected penalty given y_j and $u_{ne(j)}$.

Similar to the specification of Algorithm 2 by employing Proposition 3.3 in Algorithm 1, it is possible to obtain an iterative scheme which, starting with an initial two-stage strategy, converges to a person–by–person optimal one. The treatment of the RHSs of Eq.s (45), (47)-(53) as operators that can act on any set of their arguments, not necessarily optimal, is summarized by r_j and q_j together with f_j, g_j and h_j given by

$$\alpha_j = r_j(C_{ne(j)\to j}) \tag{54}$$

$$\beta_j = q_j(P_j) \tag{55}$$

$$P_j = f_j(P_{ne(j)\to j}) \tag{56}$$

$$P_{j\to ne(j)} = g_j(\alpha_j) \tag{57}$$

$$C_{j\to ne(j)} = h_j(\beta_j, P_{ne(j)\to j}) \tag{58}$$

where $P_{ne(j)\to j} = \{P_{i\to j}\}_{i\in ne(j)}$, $C_{ne(j)\to j} = \{C_{i\to j}\}_{i\in ne(j)}$ and $C_{j\to ne(j)} = \{C_{j\to i}\}_{i\in ne(j)}$. The resulting iterative scheme after deploying the operators given by Eq.s(54)–(58) is given by Algorithm 3.

Finally, the objective value at the l^{th} step is easily found to be

$$J(\gamma^l) = \sum_{i\in\mathcal{V}} G_i^d(\nu_i^l) + \lambda \sum_{i\in\mathcal{V}} G_i^c(\mu_i^l) \tag{59}$$

where

$$G_i^d(\nu_i^l) = \sum_{u_{ne(i)}} \int_{\mathcal{X}_i} \mathrm{d}x_i p(x_i) P_i^{l+1}(u_{ne(i)}|x_i) I_i(u_{ne(i)}, x_i; \nu_i^l) \tag{60}$$

and

$$G_i^c(\mu_i^l) = \sum_{u_i} \int_{\mathcal{X}_i} \mathrm{d}x_i c_i^c(u_i, x_i) p(x_i) p(u_i|x_i; \mu_i^l) \tag{61}$$

in terms of the expressions discussed above.

Note that, similar to that for optimizing in–network strategies over DAGs, the `Update` step of Algorithm 3 also admits a message passing interpretation. In the first pass, all nodes compute and send forward likelihood terms to their neighbors. In the second pass, upon receiption of the likelihood messages, all nodes update their stage–two estimation rules and compute and send expected cost messages to their neighbors. After receiving cost messages from neighbors, each node update its stage–one communication rule. This structure of the optimization scheme renders it suitable for a possible network self–organization requirement similar to Algorithm 2.

IV. MC Optimization Framework for in-network processing strategies over DAGs

In Section III-A1 and III-A2 we have provided conditions of optimality in a person–by–person sense rendering Algorithm 2 for the *offline* optimization of the class of decentralized estimation strategies of concern. Specifically, provided that Assumptions 1–4 hold, the operator representations d_j, f_j, g_j and h_j given by Eq.s(26)–(29) summarize Eq.s (21)-(25) respectively applied to local rules not necessarily optimal. If, in addition, Assumption 5 holds, the structures exhibited in Corollaries 3.4 and 3.6 are induced on the operators. However, it is not possible to evaluate the right hand side (RHS) of these equations and correspondingly d_j, f_j, g_j and h_j exactly, in general, for arbitrary prior marginals $p(x_j)$, observation likelihoods $p(y_j|x_j)$ and rules local to nodes other than j, i.e., $\gamma_{\setminus j}$. A similar problem arises in message passing algorithms over continous Markov random fields and has been the motivation for algorithms relying on particle representations together with approximate computational schemes including Non-parametric Belief Propagation [38], [39] which has been successfully applied in a number of contexts including articulated visual object tracking [40], [41].

In this section, we propose particle based representations together with approximate computational schemes so that Algorithm 2 can be realized. We exploit the Monte Carlo method [42], [43] and Importance Sampling [44],

[45] such that independent samples generated from only the marginal distributions of X and Y are required, i.e.,

$$S_{x_j} \triangleq \{x_j^{(1)}, x_j^{(2)}, ..., x_j^{(M_j)}\} \text{ such that } x_j^{(m)} \sim p(x_j) \text{ for } m = 1, 2, ..., M_j \tag{62}$$

and

$$S_{y_j} \triangleq \{y_j^{(1)}, y_j^{(2)}, ..., y_j^{(P_j)}\} \text{ such that } y_j^{(p)} \sim p(y_j) \text{ for } p = 1, 2, ..., P_j \tag{63}$$

for $j \in \mathcal{V}$. Although the sizes of these sets might vary for each $j \in \mathcal{V}$, we assume that $M_j = M$ and $P_j = P$ for $j \in \mathcal{V}$ for simplicity of the discussion throughout.

Generating independent samples provides scalability in the number of variables N and the number of samples M together with ease of application for a number of reasons. First, considering a single random variable, it is a relatively straightforward task to generate pseudo random numbers from an arbitrary probability density function provided that the inverse of the corresponding cumulative distribution can be evaluated (see, e.g., Chp. 2 in [45]). In addition, the neccessary knowledge of distributions to utilize Algorithm 2, i.e., $p(x_{\pi(i)}, x_i)$ and $p(y_i|x_i)$ for all $i \in \mathcal{V}$, implies that the marginals are already known and hence we do not require the knowledge of any additional distributions. Besides, we consider independent generations that require no coordinations. For the case in which we consider scalability with the number of random variables involved, sampling from the joint distribution is cumbersome where scalability can be maintained up to a degree with coordinated generation schemes, which require the evaluation of characterizing densities such as the conditionals. For example Gibbs sampling introduced in [46] requires the so called full conditionals $\{p(x_j|x_{\setminus j})\}_{j\in\mathcal{V}}$ whereas the Substitution Sampling method requires $N(N-1)$ conditionals for N components [47].

We proceed by considering the sufficient condition of person-by-person optimality for the j^{th} rule given by Proposition 3.3. The Monte Carlo optimization algorithm we propose follows successive approximations to the expressions constituting the j^{th} pbp optimal local rule (see Eq.s(17) and (20)). In Section IV-A we approximate the pbp optimal rule assuming that the factors in the RHS of Eq.(20) are known over their entire domain sets. In the second step we proceed with approximating to the incoming message likelihood (Sec.IV-B). In Section IV-C, the node–to–node terms, i.e., $P^*_{i\to j}$ and $C^*_{k\to j}$ for $i \in \pi(j)$ and $k \in \chi(j)$ respectively, are approximated and finally in Section IV-D all the approximations are utilized together comprising the proposed algorithm after a treatment of the approximations as operators in a similar fashion to our development in Section III-A2.

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

Given a pbp optimal strategy $\gamma^* \in \Gamma^{\mathcal{G}}$, consider the j^{th} optimal local rule given by Eq.s(17) and (20) in the case that the remaining are fixed at the optimum $\gamma_{\setminus j} = \gamma^*_{\setminus j}$. After substituting Eq.(20) in Eq.(17) we obtain

$$\gamma_j^*(Y_j, U_{\pi(j)}) = \arg \min_{u_j \times \hat{x}_j \in (\mathcal{U}_j \times \mathcal{X}_j)} R_j^*(u_j, \hat{x}_j; Y_j, U_{\pi(j)}) \tag{64}$$

where

$$R_j^*(u_j, \hat{x}_j; y_j, u_{\pi(j)}) = \int_{\mathcal{X}_j} \mathrm{d}x_j p(x_j) p(y_j|x_j) P_j^*(u_{\pi(j)}|x_j) \left[c_j(u_j, \hat{x}_j, x_j) + C_j^*(u_j, x_j)\right] \tag{65}$$

for all $u_j \in \mathcal{U}_j$, $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$, $y_j \in \mathcal{Y}_j$ and $\hat{x}_j \in \mathcal{X}_j$ where unlike the detection problem in [31], $\mathcal{X}_j$ is a denumerable set and the RHS of Eq.(65) involves an integral over $\mathcal{X}_j$. It is reasonable to assume that the observation likelihood $p(y_j|x_j)$ and the cost $c_j(u_j, \hat{x}_j, x_j)$ are known. However, the incoming message likelihood, i.e., $P_j^*(u_{\pi(j)}|x_j)$, together with the conditional cost induced on the descendants, i.e., $C_j^*(u_j, x_j)$, depend on the remaining local rules $\gamma_{\setminus j}^*$ (see Section III-A2) and do not necessarily admit closed form expressions for arbitrary $\gamma_{\setminus j} \in \Gamma_j^{\mathcal{G}}$.

Suppose that for all $x_j \in \mathcal{X}_j$, $P_j^*(u_{\pi(j)}|x_j)$ and $C_j^*(u_j, x_j)$ are known, i.e., it is possible to evaluate them over their entire domains. The integral on the RHS of Eq.(65) still prevents R_j^* to be evaluated exactly, in general. However, an approximation is possible through the classical Monte Carlo method given M independent samples generated from $p(x_j)$, i.e., S_{x_j} given by Eq.(62),

$$\tilde{R}_j^*(u_j, \hat{x}_j; y_j, y_{\pi(j)}) = \frac{1}{|S_{x_j}|} \sum_{x_j \in S_{x_j}} p(y_j|x_j) P_j^*(u_{\pi(j)}|x_j) \left[c_j(u_j, \hat{x}_j, x_j) + C_j^*(u_j, x_j)\right] \tag{66}$$

where tilde denotes an approximation, i.e., $\tilde{R}_j^*(u_j, \hat{x}_j; y_j, y_{\pi(j)}) \approx R_j^*(u_j, \hat{x}_j; y_j, y_{\pi(j)})$ over its entire domain. $\tilde{R}_j^*$ substituted in Eq.(64) in place of R_j^* corresponds to a local rule, which is an approximation to γ_j^*. Let us represent the approximation to the optimal local rule by $\tilde{\gamma_j^*}^1$ where the superscript 1 denotes that the approximation involves a single MC approximated function, then $\tilde{\gamma_j^*}^1(y_j, u_{\pi(j)}) \approx \gamma_j^*(y_j, u_{\pi(j)})$ for all $y_j \in \mathcal{Y}_j$ and for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$ with nonzero probability.

Since we have assumed that P_j^* and C_j^* are known, it is implied that they can be evaluated at $x_j \in S_{x_j}$, for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$ and $u_j \in \mathcal{U}_j$ respectively. $\tilde{R}_j^*$ substituted in Eq.(64) in place of R_j^* corresponds to a local rule, which is an approximation to γ_j^*. Let us represent the approximation to the optimal local rule by $\tilde{\gamma_j^*}^1$ where the superscript 1 denotes that the approximation involves a single MC approximated function, then $\tilde{\gamma_j^*}^1(y_j, u_{\pi(j)}) \approx \gamma_j^*(y_j, u_{\pi(j)})$ for all $y_j \in \mathcal{Y}_j$ and for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$ with nonzero probability.

Consider Corollary 3.4. The objective of minimization in the variational form of the j^{th} local rule given by Eq.(64) is separable, i.e. $R_j^*(u_j, \hat{x}_j; y_j, u_{\pi(j)}) = R_{j,d}^*(\hat{x}_j; y_j, u_{\pi(j)}) + R_{j,c}^*(u_j; y_j, u_{\pi(j)})$, under a separable cost function local to node j and yields two separate problems and corresponding rules for estimation and communication denoted by ν_j and μ_j respectively. Similarly the approximation $\tilde{R}_j^*$ given by Eq.(66) splits trivially to two approximations, i.e., $\tilde{\nu_j}^1$ and $\tilde{\mu_j}^1$.

Example 4.1: Consider Example 3.5, Eq.(66) substituted in Eq.(64) implies that the explicit solution for the quadratic estimation error given by Eq.(34) is approximated by

$$\hat{x}_j = \tilde{\nu_j}^1(y_j, u_{\pi(j)}) = \frac{\sum_{m=1}^{M} x_j^{(m)} p(y_j|x_j^{(m)}) P_j^*(u_{\pi(j)}|x_j^{(m)})}{\sum_{m=1}^{M} p(y_j|x_j^{(m)}) P_j^*(u_{\pi(j)}|x_j^{(m)})} \tag{67}$$

B. Approximating the message likelihood function

In the previous section, we proposed an approximation to the j^{th} optimal rule which requires the incoming message likelihood $P_j^*(u_{\pi(j)}|x_j)$ and the conditional expected cost $C_j^*(u_j, x_j)$ to be known at $x_j = x_j^{(m)}$ for $m = 1, 2, ..., M$, for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$ and for all $u_j \in \mathcal{U}_j$ respectively. Since it is not possible to express

these functions in closed form for an arbitrary set of local rules $\gamma_j \in \Gamma_j^{\mathcal{G}}$, in this step, we consider approximate computations of Eq.(21) and Eq.(23).

We continue the discussion by considering Eq.(21) for the case in which $\pi(j) \neq \emptyset$. Suppose that the forward node–to–node terms, i.e., $P^*_{i\to j}(u_{i\to j}|x_i)$ for $i \in \pi(j)$, are known such that we can evaluate them at $x_i = x_i^{(m)}$ where $x_i^{(m)} \in S_{x_i}$ and for all $u_{i\to j} \in \mathcal{U}_{i\to j}$. This assumption is justified by the fact that if the 1–step approximation described in Section IV-A were to be applied to the rules local to nodes $i \in \pi(j)$, then S_{x_i} would be utilized.

Next, we note that it is possible to treat the concatenation of the elements of the parent sample sets, i.e., S_{x_i} for $i \in \pi(j)$, as a sample set that is drawn from the product of distributions that generated them. In other words, consider $x_{\pi(j)}^{(m)} \triangleq (x_i^{(m)})_{i\in\pi(j)}$ for $m = 1, 2, ..., M$ where $x_i^{(m)} \in S_{x_i}$ for $i \in \pi(j)$. These elements constitute a sample set $S_{\pi(j)} \triangleq \{x_{\pi(j)}^{(m)} | x_{\pi(j)}^{(m)} = (x_i^{(m)})_{i\in\pi(j)}\}$ and it holds that $x_{\pi(j)}^{(m)} \sim \prod_{i\in\pi(j)} p(x_i)$.

This observation enables the Importance Sampling approximation (see, e.g., Chp. 3 in [45]) for P_j^* through the importance sampling distribution $\prod_{i\in\pi(j)} p(x_i)$. Then the importance weights are given by

$$\omega_j^{(m)(m')} = p(x_{\pi(j)}^{(m')}|x_j^{(m)}) / \prod_{i\in\pi(j)} p(x_i^{(m')})$$

with the corresponding approximation

$$\tilde{P_j^*}^1(u_{\pi(j)}|x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M} \omega_j^{(m)(m')}} \sum_{m'=1}^{M} \omega_j^{(m)(m')} \prod_{i\in\pi(j)} P^*_{i\to j}(u_{i\to j}|x_i^{(m')}) \tag{68}$$

for $m = 1, 2, ..., M$ and for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$.

Now let us turn to the computation of the conditional expected cost $C_j^*(u_j, x_j)$ and consider Eq.(23) for the case in which $\chi(j) \neq \emptyset$. We assume that the node–to–node backward cost terms, i.e., for all $k \in \chi(j)$, $C^*_{k\to j}(u_{j\to k}, x_j)$, are known at $x_j = x_j^{(m)}$ for $m = 1, 2, ..., M$ and for all $u_{j\to k} \in \mathcal{U}_{j\to k}$. Then, the required values, i.e., $C_j^*(u_j, x_j^{(m)})$ for $m = 1, 2, ..., M$ and for all $u_j \in \mathcal{U}_j$, can be computed exactly using Eq.(23).

From node j' s point of view, given node–to–node terms $P^*_{i\to j}$ and $C^*_{k\to j}$ evaluated at points generated from the appropriate marginal distributions, a further approximation to the j^{th} pbp optimal rule is obtained by computing $\tilde{P_j^*}^1$ and C_j^* at values of their arguments required in Eq.(66) and substituting $\tilde{P_j^*}^1$ in place of P_j^*. Let $\tilde{\gamma_j^*}^2$ denote the corresponding rule, then $\tilde{\gamma_j^*}^2(y_j, u_{\pi(j)}) \approx \tilde{\gamma_j^*}^1(y_j, u_{\pi(j)}) \approx \gamma_j^*(y_j, u_{\pi(j)})$ for all $y_j \in \mathcal{Y}_j$ and for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$ with nonzero probability.

C. Approximating the node–to–node terms

In the previous section, the approximation to the j^{th} local rule is introduced under the conditions that for all $i \in \pi(j)$, $P^*_{i\to j}(u_{i\to j}|x_i)$ is known for all $u_{i\to j} \in \mathcal{U}_{i\to j}$ and $x_i = x_i^{(m)}$ for $x_i^{(m)} \in S_{x_i}$. Another requirement is to be able to evaluate $C^*_{k\to j}(u_{j\to k}, x_j)$ for all $u_{j\to k} \in \mathcal{U}_{j\to k}$ and $x_j = x_j^{(m)}$ where $x_j^{(m)} \in S_{x_j}$. Therefore, a further step, which is of concern in this subsection, involves approximating the node-to-node terms $P^*_{i\to j}$ and $C^*_{k\to j}$ evaluated at the discretization of their domains provided by the sample sets.

We consider the parent nodes $i \in \pi(j)$ and consider evaluation of Eq.(22) at the required values of its arguments. Suppose that γ_i^* is fixed at the optimum, implying also that $p(u_i, \hat{x}_i|y_i, u_{\pi(i)}; \gamma_i^*)$ is specified through Eq.s(5) and (6) for all $i \in \pi(i)$. The multiple integral term in Eq.(22), rewritten here as

$$p(u_i|x_i, u_{\pi(i)}; \gamma_i^*) = \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i \int_{\mathcal{Y}_i} \mathrm{d}y_i \, p(u_i, \hat{x}_i|y_i, u_{\pi(i)}; \gamma_i^*) p(y_i|x_i)$$

for convenience, should be evaluated at $x_i = x_i^{(m)}$ for $m = 1, 2, ..., M$, for all $u_i \in \mathcal{U}_i$ and for all $u_{\pi(i)} \in \mathcal{U}_{\pi(i)}$. Since there is no closed form solution for arbitrary choice of γ_i^* and the likelihood $p(y_i|x_i)$, we perform an Importance Sampling approximation through the importance sampling distribution $p(y_i)$. Utilizing $y_i^{(p)} \in S_{y_i}$ and the importance weights given by

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

an importance sampling approximation to $p(u_i|x_i^{(m)}, u_{\pi(i)}; \gamma_i^*)$ for $m = 1, 2, ..., M$, for all $u_i \in \mathcal{U}_i$ and for all $u_{\pi(i)} \in \mathcal{U}_{\pi(i)}$ is given by

$$\tilde{p}(u_i|x_i^{(m)}, u_{\pi(i)}; \gamma_i^*) = \frac{1}{\sum_{p=1}^{P} \omega_i^{(m)(p)}} \sum_{p=1}^{P} \omega_i^{(m)(p)} \delta_{u_i, [\gamma_i^*(y_i^{(p)}, u_{\pi(i)})]_{\mathcal{U}_i}} \tag{69}$$

where δ denotes the Kronecker's delta. Note that, if Assumption 5 holds, the estimation and communication rules separate and the discussion above applies with $p(u_i|x_i, u_{\pi(i)}; \gamma_i^*) = p(u_i|x_i, u_{\pi(i)}; \mu_i^*)$.

Regarding Eq.(22), having approximated the multiple integral term for $j \in \mathcal{V}$, we similarly assume that $P_i^*(u_{\pi(i)}|x_i)$ is known for $i \in \pi(j)$, for $x_i = x_i^{(m)}$ such that $x_i^{(m)} \in S_{x_i}$, and for all $u_{\pi(i)} \in \mathcal{U}_{\pi(i)}$. Together with Eq.(69) we obtain

$$\tilde{P}_{i\to j}^*(u_{i\to j}|x_i^{(m)}) = \sum_{u_{\chi(i)\setminus j} \in U_{\chi(i)\setminus j}} \sum_{u_{\pi(i)} \in u_{\pi(i)}} P_i^*(u_{\pi(i)}|x_i^{(m)}) \tilde{p}(u_i|u_{\pi(i)}, x_i^{(m)}; \gamma_i^*) \tag{70}$$

for $m = 1, 2, ..., M$ and for all $u_{i\to j} \in \mathcal{U}_{i\to j}$. It is possible to replace the node–to–node terms assumed to be known in Eq.(68) with Eq.(70) and obtain a further step in the progressive approximations to γ_j^*.

The remaining term to consider is the conditional expected costs induced on the descendants of j on the branch initiating with its child k, i.e., $C_{k\to j}^*(u_{j\to k}, x_j)$, for all $k \in \chi(j)$, evaluated at $x_j = x_j^{(m)}$ where $x_j^{(m)} \in S_{x_j}$ and for all $u_{j\to k} \in \mathcal{U}_{j\to k}$. A similar reasoning leads to approximating the required values through utilizing Monte Carlo methods on the RHS of the expression obtained by substituting Eq.(25) in Eq.(24).

Consider Eq.(25) and suppose that γ_k^* is known for any $k \in \chi(j)$ also implying that $p(u_k, \hat{x}_k|y_k, u_{\pi(k)}; \gamma_k^*)$ is determined. Substituting Eq.(5) and (6) in Eq.(25) yields

$$\begin{aligned} I^*(u_{\pi(k)}, x_k; \gamma_k^*) = \int_{\mathcal{Y}_k} \mathrm{d}y_k \, [\; & c_k(\, [\gamma_k^*(y_k, u_{\pi(k)})]_{\mathcal{U}_k}, \, [\gamma_k^*(y_k, u_{\pi(k)})]_{\mathcal{X}_k}, \, x_k) \\ & + C_k^*(\, [\gamma_k^*(y_k, u_{\pi(k)})]_{\mathcal{U}_k}, \, x_k) \;] p(y_k|x_k) \end{aligned} \tag{71}$$

evaluation of which can be approximated at $x_k = x_k^{(m)}$ for all $x_k^{(m)} \in S_{x_k}$ and for all $u_{\pi(k)} \in \mathcal{U}_{\pi(k)}$ by the Importance Sampling method, using the importance density $p(y_k)$. Assuming $C_k^*(u_k, x_k)$ is known at $x_k = x_k^{(m)}$

where $x_k^{(m)}$ is an element of the usual sample set local to k, i.e., S_{x_k}, and for all $u_k \in \mathcal{U}_k$ and utilizing $y_k^{(p)} \in S_{y_k}$ together with the importance weights

$$\omega_k^{(m)(p)} = p(y_k^{(p)}|x_k^{(m)})/p(y_k^{(p)})$$

we obtain

$$\tilde{I^*}(u_{\pi(k)}, x_k^{(m)}; \gamma_k^*) = \frac{1}{\sum_{p=1}^{P} \omega_k^{(m)(p)}} \sum_{p=1}^{P} \omega_k^{(m)(p)} [\ c_k(\ [\gamma_k^*(y_k^{(p)}, u_{\pi(k)})]_{\mathcal{U}_k},\ [\gamma_k^*(y_k^{(p)}, u_{\pi(k)})]_{\mathcal{X}_k},\ x_k) + C_k^*(\ [\gamma_k^*(y_k^{(p)}, u_{\pi(k)})]_{\mathcal{U}_k},\ x_k)\ ] \quad (72)$$

for $m = 1, 2, ..., M$ and for all $u_{\pi(k)} \in \mathcal{U}_{\pi(k)}$ such that $\tilde{I^*}(u_{\pi(k)}, x_k^{(m)}; \gamma_k^*) \approx I^*(u_{\pi(k)}, x_k^{(m)}; \gamma_k^*)$ holds [8].

In addition, if Assumption 5 holds, we consider Corollary 3.6 yielding the Importance Sampling approximations to Eq.(36) and Eq.(37) evaluated at $x_k^{(m)}$ and $u_{\pi(k)}$ to be similarly obtained as

$$\begin{aligned} \tilde{J}_{d|x_k^{(m)}, u_{\pi(k)}} &= \frac{1}{\sum_{p=1}^{P} \omega_k^{(m)(p)}} \sum_{p=1}^{P} \omega_k^{(m)(p)} c_k^d(\nu_k^*(y_k^{(p)}, u_{\pi(k)}), x_k^{(m)}) \\ \tilde{J}_{c|x_k^{(m)}, u_{\pi(k)}} &= \sum_{u_k} \left(\lambda c_k^c(u_k, x_k^{(m)}) + C_k^*(u_k, x_k^{(m)}) \right) \tilde{p}(u_k|x_k^{(m)}, u_{\pi(k)}; \mu_k^*) \end{aligned}$$

respectively. Then $\tilde{I^*}(u_{\pi(k)}, x_k^{(m)}; \gamma_k^*) = \tilde{J}_{d|x_k^{(m)}, u_{\pi(k)}} + \tilde{J}_{c|x_k^{(m)}, u_{\pi(k)}}$ holds.

Eq.(24) requires message likelihood terms from all parents of node k except node j and it is reasonable to assume that for any $j' \in \pi(k) \setminus j$, $P_{j' \to k}^*(u_{j' \to k}|x_j')$ is known at $x_j' = x_{j'}^{(m)}$ for $x_{j'}^{(m)} \in S_{x_{j'}}$ and for all $u_{j' \to k} \in \mathcal{U}_{j' \to k}$. Similarly, we observe that the set which is constituted of elements that are concetanation of elements from the usual sample sets local to $j' \in \pi(k) \backslash j$ is distributed according to the product of the corresponding marginals. In other words, let us define $x_{\pi(k)\backslash j}^{(m)} \triangleq (x_{j'}^{(m)})_{j' \in \pi(k)\backslash j}$. Then it holds that $x_{\pi(k)\backslash j}^{(m)} \sim \prod_{j' \in \pi(k)\backslash j} p(x_{j'})$ and an importance sampling approximation to Eq.(24) is possible through the importance distribution $\prod_{j' \in \pi(k)\backslash j} p(x_{j'})$. Having computed $\tilde{I^*}(u_{\pi(k)}, x_k^{(m)}; \gamma_k^*)$ and utilizing the usual sample sets local to nodes $j' \in \pi(k) \backslash j$ together with the importance sampling weigths

$$\omega^{(m)(m')} = p(x_{\pi(k)\backslash j}^{(m')}, x_k^{(m')}|x_j^{(m)})/p(x_k^{(m')}) \prod_{j' \in \pi(k)\backslash j} p(x_{j'}^{(m')})$$

we obtain

$$\tilde{C}_{k \to j}^*(u_{j \to k}, x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M} \omega^{(m)(m')}} \sum_{m'=1}^{M} \omega^{(m)(m')} \sum_{u_{\pi(k)\backslash j}} \prod_{j' \in \pi(k)\backslash j} P_{j' \to k}^*(u_{j' \to k}|x_{j'}^{(m')}) \tilde{I^*}(u_{\pi(k)}, x_k^{(m')}; \gamma_k^*) \quad (73)$$

for $m = 1, 2, ..., M$ and for all $u_{j \to k} \in \mathcal{U}_{j \to k}$ which[9], after substituting in place of $C_{k \to j}^*$ in the RHS of Eq.(23)

[8]Note that $[\gamma_k^*(y_k^{(p)}, u_{\pi(k)})]_{\mathcal{U}_k}$ and $[\gamma_k^*(y_k^{(p)}, u_{\pi(k)})]_{\mathcal{X}_k}$ are simply the communication symbol and estimation output of γ_k^* evaluated at the tuple $(y_k^{(p)}, u_{\pi(k)})$.

[9]Note that we have approximated the forward likelihood terms regarding node j and its parents, i.e. $P_{i \to j}^*$ for $i \in \pi(j)$. However, we still assume that node-to-node terms regarding other nodes including $P_{j' \to k}^*$ for $j' \in \pi(k) \backslash j$ where $k \in \chi(j)$ are known over all their domains.

for $\chi(j) \neq \emptyset$ yields $\tilde{C}^*_j$, i.e.,

$$\tilde{C}^*_j(u_j, x_j^{(m)}) = \sum_{k \in \chi(j)} \tilde{C}^*_{k \to j}(u_{j \to k}, x_j^{(m)}) \tag{74}$$

for $m = 1, 2, ..., M$ and for all $u_j \in \mathcal{U}_j$.

As a result, after substituting $\tilde{P}^*_{i \to j}$ in place of $P^*_{i \to j}$ in the RHS of Eq.(68), we obtain a further approximation to P^*_j given by

$$\tilde{P}_j^{*\,2}(u_{\pi(j)}|x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M} \omega_j^{(m)(m')}} \sum_{m'=1}^{M} \omega_j^{(m)(m')} \prod_{i \in \pi(j)} \tilde{P}^*_{i \to j}(u_{i \to j}|x_i^{(m')}) \tag{75}$$

for $m = 1, 2, ..., M$ and for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$. This approximation together with $\tilde{C}^*_{k \to j}$ given by Eq.(74) employed in $\tilde{R}^*_j$ yields $\tilde{\gamma}_j^{*\,3}(y_j, u_{\pi(j)}) \approx \tilde{\gamma}_j^{*\,2}(y_j, u_{\pi(j)}) \approx \tilde{\gamma}_j^{*\,1}(y_j, u_{\pi(j)}) \approx \gamma^*_j(y_j, u_{\pi(j)})$ for all $y_j \in \mathcal{Y}_j$ and for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$ with nonzero probability.

D. MC Optimization of in–network processing strategies over DAGs

In Section IV-A–IV-C we have introduced a Monte Carlo approximation framework regarding the sufficient conditions of person–by–person optimality given in Proposition 3.3. Considering a pbp optimal decentralized estimation strategy constrained by a polytree $\mathcal{G}$, i.e., $\gamma^* \in \Gamma^{\mathcal{G}}$ and having $\gamma_{\setminus j}$ fixed at the corresponding set of optimal local rules, i.e. $\gamma_{\setminus j} = \gamma^*_{\setminus j}$, we have constructed a rule local for j, $\tilde{\gamma}_j^{*\,3}(y_j, u_{\pi(j)})$ such that it is an approximation to the optimal rule γ^*_j given by Eq.(17) following the progression

$$\tilde{\gamma}_j^{*\,1}(y_j, u_{\pi(j)}) = \arg \min_{u_j \times \hat{x}_j \in (\mathcal{U}_j \times \mathcal{X}_j)} \sum_{x_j \in S_{x_j}} p(y_j|x_j) P^*_j(u_{\pi(j)}|x_j) \left[c_j(u_j, \hat{x}_j, x_j) + C^*_j(u_j, x_j) \right]$$

where S_{x_j} is given by Eq.(62),

$$\tilde{\gamma}_j^{*\,2}(y_j, u_{\pi(j)}) = \arg \min_{u_j \times \hat{x}_j \in (\mathcal{U}_j \times \mathcal{X}_j)} \sum_{x_j \in S_{x_j}} p(y_j|x_j) \tilde{P}_j^{*\,1}(u_{\pi(j)}|x_j) \left[c_j(u_j, \hat{x}_j, x_j) + C^*_j(u_j, x_j) \right]$$

where $\tilde{P}_j^{*\,1}$ is given by Eq.(68),

$$\tilde{\gamma}_j^{*\,3}(y_j, u_{\pi(j)}) = \arg \min_{u_j \times \hat{x}_j \in (\mathcal{U}_j \times \mathcal{X}_j)} \sum_{x_j \in S_{x_j}} p(y_j|x_j) \tilde{P}_j^{*\,2}(u_{\pi(j)}|x_j) \left[c_j(u_j, \hat{x}_j, x_j) + \tilde{C}^*_j(u_j, x_j) \right] \tag{76}$$

where $\tilde{C}^*_j(u_j, x_j)$ and $\tilde{P}_j^{*\,2}$ are given by Eq.s (74) and (75) respectively. Hence, in order to obtain $\tilde{\gamma}_j^{*\,3}$ we have utilized the proposed particle representations and approximate computational schemes for all terms that depend on $\gamma^*_{\setminus j}$ including the node–to–node terms. Note that, we have not approximated $\gamma^*_{\setminus j}$ up to this point and assumed that it is known exactly.

On the other hand, given S_{x_j} and S_{y_j}, the approximation framework is valid for the rules local to any node $j \in \mathcal{V}$: Owing to fusing the forward message likelihoods via importance sampling, the node–to–node terms given by Eq.s(70) and (73) utilize the discretization provided by these sets regardless of which node' s local rule is subject to approximation. Hence, it is possible to treat the RHS of the expressions within the framework as operators valid for any strategy $\gamma \in \Gamma^{\mathcal{G}}$ including those in the "approximating" form given by Eq. (76). For the rest of this paper,

an approximation to a function that appears in the local rules refers to its corresponding approximation in Eq.(76) and we denote these functions without any further superscripts, e.g., we denote $\tilde{\gamma_j^*}^3$ with $\tilde{\gamma_j^*}$. Let us summarize the Monte Carlo framework with

$$\begin{aligned}
\tilde{\phi_j}(S_{x_j}, \hat{x}_j) &= \tilde{d_j}(\tilde{P_j}(S_{x_j}), \tilde{C}_{\chi(j)\to j}) \\
\tilde{P_j}(S_{x_j}) &= \tilde{f_j}(\tilde{P}_{\pi(j)\to j}) \\
\tilde{P}_{j\to\chi(j)} &= \tilde{g_j}(\tilde{\phi_j}(S_{x_j}, \hat{x}_j), \tilde{P_j}(S_{x_j})) \\
\tilde{C}_{j\to\pi(j)} &= \tilde{h_j}(\tilde{\phi_j}(S_{x_j}, \hat{x}_j), \tilde{P}_{\pi(j)\to j}, \tilde{C}_{\chi(j)\to j})
\end{aligned}$$

where

$$\tilde{P_j}(S_{x_j}) = \{(\tilde{P_j}(u_{\pi(j)}|x_j), u_{\pi(j)}, x_j) | u_{\pi(j)} \in \mathcal{U}_{\pi(j)} \wedge x_j \in S_{x_j}\}$$

$$\tilde{P}_{\pi(j)\to j} = \{\tilde{P}_{i\to j}(S_{x_i})\}_{i\in\pi(j)}$$

$$\tilde{P}_{i\to j}(S_{x_i}) = \{(\tilde{P}_{i\to j}(u_{i\to j}, x_i), u_{i\to j}, x_i) | u_{i\to j} \in \mathcal{U}_{i\to j} \wedge x_i \in S_{x_i}\}$$

$$\tilde{P}_{j\to\chi(j)} = \{\tilde{P}_{j\to k}(S_{x_j})\}_{k\in\chi(j)}$$

$$\tilde{C}_{\chi(j)\to j} = \{\tilde{C}_{k\to j}(S_{x_j})\}_{k\in\chi(j)}$$

and $\tilde{\phi_j}(S_{x_j}, \hat{x}_j)$ is given by

$$\left\{(p(y_j|x_j)\tilde{P_j}(u_{\pi(j)}|x_j)\left[c(u_j, \hat{x}_j, x_j) + \tilde{C_j}(u_j, x_j)\right], u_j, x_j) | u_j \in \mathcal{U}_j, u_{\pi(j)} \in \mathcal{U}_{\pi(j)}, x_j \in S_{x_j}\right\}$$

Note that $\tilde{C}_{k\to j}(S_{x_j})$ implies a definition in a similar fashion to that for $\tilde{P}_{i\to j}(S_{x_i})$. Note also that $\tilde{\phi_j}(S_{x_j}, \hat{x}_j)$ is not a complete discretization of ϕ_j, i.e., considering Eq.(20), for the evaluation of $\phi_j(u_j, \hat{x}_j, x_j; u_{\pi(j)})$, the argument $\hat{x}_j$ needs not to be discretized since only $c(u_j, \hat{x}_j, x_j)$ acts on it and it is assumed to be known over its entire domain. Therefore $\hat{x}_j$ is a free variable that can take values from $\mathcal{X}_j$. On the other hand, the conventional Monte Carlo approximation drops $p(x_j)$ and discretizes ϕ in x_j.

It is immediately possible to employ this framework in Algorithm 2 and achieve a Monte Carlo optimization algorithm which, starting with initial local rules, iteratively results in a strategy that corresponds to performing computations to approximate a person-by-person optimal one. Given by Algorithm 4, this scheme maintains the message passing interpretation appearing in the `Update` step of Algorithm 2.

Starting with $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\{\mathcal{U}_{i\to j} | (i,j) \in \mathcal{E}\}$, each node initially maintains the knowledge of $p(x_{\pi(j)}, x_j)$ and $c(u_j, \hat{x}_j, x_j)$. As soon as samples from the marginal distributions, i.e., S_{x_j}, together with samples from the marginal distributions of the observation processes, i.e., S_{y_j}, are generated for all $j \in \mathcal{V}$, and an initial local rule $\gamma_j^0 \in \Gamma_j^{\mathcal{G}}$ is selected, the iterative scheme yields a set of local rules such that each node performs computations corrresponding to an approximation to a person–by–person optimum.

The approximate computation of the expected cost required in the `Check` step of Algorithm 4 for any given strategy, i.e., $\tilde{J}(\gamma)$ is obtained through a Monte Carlo approximation $\tilde{G}_j(\gamma_j^l)$ to Eq.(42) using the usual sample sets,

Algorithm 4 Iterations converging to an approximate pbp optimal decentralized estimation strategy over a DAG $\mathcal{G}$.

1: Choose $\gamma^0 = (\gamma_1^0, \gamma_2^0, ..., \gamma_N^0)$ such that $\gamma_j^0 \in \Gamma_j^{\mathcal{G}}$ for $j = 1, 2, ..., N$; Choose $\varepsilon \in \mathbb{R}^+$;$l = 0$ ▷ `Initialize`

2: $l = l + 1$

3: `For` $j = 1, 2, ..., N$ `Do` ▷ `Update Step 1: Forward Pass`

$$\tilde{P_j}^l(S_{x_j}) = \tilde{f}_j(\left\{\tilde{P}^l_{i\to j}(S_{x_i})\right\}_{i\in\pi(j)})$$

$$\left\{\tilde{P}^l_{j\to k}(S_{x_j})\right\}_{k\in\chi(j)} = \tilde{g}_j(\tilde{\phi_j}^{l-1}(S_{x_j}, \hat{x}_j), \tilde{P}^l{}_j(S_{x_j}))$$

4: `For` $j = N, N-1, ..., 1$ `Do` ▷ `Update Step 2: Backward Pass`

$$\tilde{\phi_j}^l(S_{x_j}, \hat{x}_j) = \tilde{d}_j(\tilde{P_j}^l(S_{x_j}), \left\{\tilde{C}^l_{k\to j}(S_{x_j})\right\}_{k\in\chi(j)})$$

$$\left\{\tilde{C}^l_{j\to i}(S_{x_i})\right\}_{i\in\pi(j)} = \tilde{h}_j(\tilde{\phi_j}^l(S_{x_j}, \hat{x}_j), \left\{\tilde{P}^l_{i\to j}(S_{x_i})\right\}_{i\in\pi(j)}, \left\{\tilde{C}^l_{k\to j}(S_{x_j})\right\}_{k\in\chi(j)})$$

5: `If` $\tau(\tilde{J}(\tilde{\gamma}^l), \tilde{J}(\tilde{\gamma}^{l-1}), ..., \tilde{J}(\tilde{\gamma}^0)) < \varepsilon$ `STOP, else GO TO` 2 ▷ `Check`

i.e., S_{x_j} and S_{y_j}, as

$$\tilde{G}_j(\tilde{\gamma}_j^l) = \frac{1}{M}\sum_{m=1}^{M}\sum_{u_{\pi(j)}\in\mathcal{U}_{\pi(j)}} \tilde{P}_j^{l+1}(u_{\pi(j)}|x_j^{(m)})\times \frac{1}{\sum_{p=1}^{P}\omega_k^{(m)(p)}}\sum_{p=1}^{P}\omega_k^{(m)(p)} c_j(\,[\gamma_j(y_j^{(p)}, u_{\pi(j)})]_{\mathcal{U}_j},\,[\gamma_j(y_j^{(p)}, u_{\pi(j)})]_{\mathcal{X}_j}, x_j^{(m)}) \quad (77)$$

where $\omega_k^{(m)(p)} = p(y_k^{(p)}|x_k^{(m)})/p(y_k^{(p)})$. If Assumption 5 holds, the expression above turns to

$$\tilde{G}_j(\tilde{\gamma}_j^l) = \frac{1}{M}\sum_{m=1}^{M}\sum_{u_{\pi(j)}\in\mathcal{U}_{\pi(j)}} \tilde{P}_j^{l+1}(u_{\pi(j)}|x_j^{(m)})\left[\tilde{J}_{d|x_j^{(m)},u_{\pi(j)}} + \lambda\sum_{u_j\in\mathcal{U}_j} c_j^c(u_j, x_j^{(m)})\tilde{p}(u_j|x_j^{(m)}, u_{\pi(j)}; \mu_j^l)\right] \quad (78)$$

and after distributing the multiplication in the RHS of the equation above and substituting in Eq.(41) in place of $G_j(\gamma_j^l)$, we obtain $\tilde{J}(\tilde{\gamma}) = \tilde{J}_d(\tilde{\gamma}) + \lambda\tilde{J}_c(\tilde{\gamma})$.

Note that $\{J(\gamma^l)|l = 0, 1, 2, ...\}$ obtained through Algorithm 2 is non–increasing whereas $\{\tilde{J}(\tilde{\gamma}^l)\}$ in Algorithm 4, being a Monte Carlo approximation to the former, does not necessarily exhibit this property. Let us define an approximation error sequence $err[l] = J(\gamma^l) - \tilde{J}(\tilde{\gamma}^l)$. This sequence will be identically zero with probability one as $M, P \to \infty$. For finite M and P, it is possible to smooth the fluctuation of $err[l]$ through filtering and utilize the corresponding termination condition, e.g., check whether $\tilde{J}(\tilde{\gamma}^l) * h[l] < \varepsilon$ where $h[l]$ is the impulse response of a linear, time invariant filter and $*$ denotes convolution. In general, a sequence that is non–increasing with high probability can be obtained through an operator τ (`Check` step of Algorithm 4), investigation of which is beyond the scope of this work.

V. MC Optimization Framework for two-Stage in-network processing strategies over UGs

In Section III-B, we presented the structure of person–by–person optimal local rules for two–stage in-network processing strategies over undirected graphs provided that certain assumptions hold. Specifically, the j^{th} local rule pair constituted of the stage–one communication and stage–two estimation rules is given by Proposition 3.7 under Assumptions 1-3 and 5. The operator representations r_j, q_j, f_j, g_j and h_j given by given by Eq.s (54)–(58) summarize Eq.s (45), (47)-(53) respectively and can be applied to arbitrary local rules not necessarily optimal. The steps of Algorithm 3 involve these operators and hence it is not possible to carry out them exactly, in general, similar to the DAG case. We similarly employ particle representations and approximate computational schemes through Monte Carlo methods in accordance with Proposition 3.7.

In Algorithm 3, each node $j \in \mathcal{V}$ starts with the knowledge of $p(x_{ne(j)}, x_j)$ and $p(y_j|x_j)$ together with an initial local rule $\gamma_j^0 \in \Gamma_j^{\mathcal{G}}$. We consider the sample sets given by Eq.(62) and (63) for $j \in \mathcal{V}$ and assume that $M_j = M$ and $P_j = P$ for simplicity. Similar to the discussion in Section IV, we approximate the expressions involved in the j^{th} pbp optimal local rule given in Proposition 3.7 in a progressive fashion. In Section V-A we approximate to the local rule pair under the assumption that both α_j^* and β_j^* are known. In the next step, we approximate β_j^* through the incoming message likelihood function (Sec. V-B) and then proceed with the computations of the node–to–node terms at the sample points and obtain further approximations to both β_j^* and α_j^* (Sec. V-C). Finally in Section V-D we employ all the previous steps simultaneously in Algorithm 3 and obtain a Monte Carlo optimization scheme which scales with both the sample sizes and the number of nodes. In addition, the message passing structure in the `Update` step of Algorithm 3 together with the amenability for network self-organization are also preserved.

A. Approximating the person–by–person optimal local rule

Consider Proposition 3.7 and the j^{th} person–by–person optimal local rule pair of stage–one communication and stage–two estimation rules given by Eq.s(44),(45) and (46),(47) respectively. Suppose that both α_j^* given in Eq.(45) and β_j^* given in Eq. (47) are known over their entire domains. Although it is relatively reasonable to assume that we are able to evaluate $p(y_j|x_j)$, $c_j^c(u_j, x_j)$ and $c_j^d(\hat{x}_j, x_j)$ for their entire domains, the incoming message likelihood and the conditional expected cost, i.e., $P_j^*(u_{ne(j)}|x_j)$ and $C_j^*(u_j, x_j)$ depend on the remaining local rule pairs and do not necessarily lead to tractable forms for arbitrary $\gamma_{\setminus j} \in \Gamma_j^{\mathcal{G}}$. Moreover, the local rules given in Eq.(44) and Eq.(46) are in a variational form such that the costs require integrations over $\mathcal{X}_j$ and hence, it is not possible to evaluate them exactly.

In the first step, we approximate these costs through the conventional Monte Carlo method with the aforementioned assumption that all the integrands are known over their entire domains. Given the usual sample set S_{x_j} as defined in Eq.(62) and considering Eq.(44) the Monte Carlo method yields

$$\tilde{\mu_j^*}^1(y_j) = \arg\min_{u_j \in \mathcal{U}_j} \sum_{x_j \in S_{x_j}} p(y_j|x_j)[\lambda c_j^c(u_j, x_j) + C_j^*(u_j, x_j)] \tag{79}$$

as an approximation to the stage–one communication rule for all $y_j \in \mathcal{Y}_j$ with non-zero probability, i.e., $\tilde{\mu_j^*}^1(y_j) \approx \mu_j^*(y_j)$ where the superscript 1 denotes that the expression involves a single MC approximation.

Considering the local estimation rule given by Eq.(46) the Monte Carlo method yields

$$\tilde{\nu_j^*}^1(y_j, u_{ne(j)}) = \arg\min_{\hat{x}_j \in \mathcal{X}_j} \sum_{x_j \in S_{x_j}} p(y_j|x_j) P_j^*(u_{ne(j)}|x_j) c_j^d(\hat{x}_j, x_j) \tag{80}$$

for all $y_j \in \mathcal{Y}_j$ and $u_{ne(j)} \in \mathcal{U}_{ne(j)}$ with non-zero probability such that $\tilde{\nu_j^*}^1(y_j, u_{ne(j)}) \approx \nu_j^*(y_j, u_{ne(j)})$.

Example 5.1: 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 one–step approximation to the j^{th} person–by–person optimal estimation rule given by Eq.(80) yields

$$\tilde{\nu_j^*}^1(y_j, u_{ne(j)}) = \frac{\sum_{m=1}^{M} x_j^{(m)} p(y_j|x_j^{(m)}) P_j^*(u_{ne(j)}|x_j^{(m)})}{\sum_{m=1}^{M} p(y_j|x_j^{(m)}) P_j^*(u_{ne(j)}|x_j^{(m)})}$$

B. Approximating the message likelihood function

The one–step approximation to the estimation rule local to node j (Eq.(80)) requires that the message likelihood function, i.e., $P_j^*(u_{ne(j)}|x_j)$ is known at $x_j = x_j^{(m)}$ for $m = 1, 2, ..., M$ and for all $u_{ne(j)} \in \mathcal{U}_{ne(j)}$. Since the RHS of Eq.(48) do not lead tractable expressions, in general, for arbitrary choices of $\gamma_{\setminus j} \in \Gamma_j^{\mathcal{G}}$, (considering the recursion involving Eq.s(49) and (50)) we approximate to P_j^* in this step.

Suppose that the node–to–node terms from the neighbors, i.e., $P_{i\to j}^*(u_{i\to j}, x_i)$ for $i \in ne(j)$, are known at $x_i = x_i^{(m)}$ where $x_i^{(m)} \in S_{x_i}$ and for all $u_{i\to j} \in \mathcal{U}_{i\to j}$. Note that if the one–step approximations to the rule pairs for the other than the j^{the} are employed, then S_{x_i}s would be employed.

We consider Eq.(48) and costruct a new sample set whose m^{th} element is the vector obtained by concatenating m^{th} elements from S_{x_i} for all $i \in ne(j)$, i.e., $S_{ne(j)} \triangleq \{x_{ne(j)}^{(m)} | x_{ne(j)}^{(m)} = (x_i^{(m)})_{i\in ne(j)}\}$. Note that $x_{ne(j)}^{(m)} \sim \prod_{i\in ne(j)} p(x_i)$ and an Importance Sampling approximation to $P_j^*(u_{ne(j)}|x_j^{(m)})$ is possible utilizing the importance sampling density $\prod_{i\in ne(j)} p(x_i)$ with the importance weights

$$\omega_j^{(m)(m')} = \frac{p(x_{ne(j)}^{(m')}|x_j^{(m)})}{\prod_{i\in ne(j)} p(x_i^{(m')})}$$

as

$$\tilde{P_j^*}^1(u_{ne(j)}|x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M} \omega_j^{(m)(m')}} \sum_{m'=1}^{M} \omega_j^{(m)(m')} \prod_{i\in ne(j)} P_{i\to j}^*(u_{i\to j}|x_i^{(m')}) \tag{81}$$

such that $\tilde{P_j^*}^1(u_{ne(j)}|x_j^{(m)}) \approx P_j^*(u_{ne(j)}|x_j^{(m)})$ for all $u_{ne(j)} \in \mathcal{U}_{ne(j)}$ and for all $x_j^{(m)} \in S_{x_j}$.

After replacing P_j^* with $\tilde{P_j^*}^1$ in Eq.(80), we obtain a further step approximation (two–steps approximation) for the person-by-person optimal estimation rule local to node j, i.e.,

$$\tilde{\nu_j^*}^2(y_j, u_{ne(j)}) \approx \tilde{\nu_j^*}^1(y_j, u_{ne(j)}) \approx \nu_j^*(y_j, u_{ne(j)})$$

for all $y_j \in \mathcal{Y}_j$ and $u_{ne(j)} \in \mathcal{U}_{ne(j)}$ with non-zero probability provided that the node-to-node likelihood terms $P_{i\to j}^*$ for all $i \in ne(j)$ are known at the required points of its domain.

Similarly, considering Eq.(79), the evaluation of C_j^* given by Eq.(51) is required, and at this point, it is convenient to assume that the node–to–node conditional cost terms, i.e., $C^*_{i\to j}(u_{j\to i}, x_j)$ for $i \in ne(j)$, at $u_{j\to i} \in \mathcal{U}_{j\to i}$ and $x_j = x_j^{(m)}$ where $x_j^{(m)} \in S_{x_j}$, are known leading to an exact evaluation of $\tilde{\mu_j^*}^1$.

C. Approximating the node–to–node terms

In the previous section, an approximation to the local estimation rule $\tilde{\nu_j^*}^2(y_j, u_{ne(j)})$ is constructed under the conditions that the message likelihood terms $P^*_{i\to j}(u_{i\to j}, x_i)$ from all neighbor nodes $i \in ne(j)$, at $x_i = x_i^{(m)}$ where $x_i^{(m)} \in S_{x_i}$ and for all $u_{i\to j} \in \mathcal{U}_{i\to j}$. Similarly, the 1–step approximation to the local communicatin rule given by Eq.(79) requires the expected cost terms $C^*_{i\to j}(u_{j\to i}, x_j)$ for $i \in ne(j)$, at $u_{j\to i} \in \mathcal{U}_{j\to i}$ and $x_j = x_j^{(m)}$ where $x_j^{(m)} \in S_{x_j}$. In this Section, we approximate these node–to–node terms and obtain a further step approximation to the local rule pair.

First, we note that Eq.(49) and note that $P^*_{i\to j}$ is a marginalization of $p(u_i|x_i;\mu_i^*)$. Then, we consider Eq.(50) and the assumption of Proposition 3.7 that all the rules local to nodes other than j^{th} are fixed at the optimum yielding $\mu_i = \mu_i^*$ for all $i \in ne(j)$. Also considering Eq.(7), it is possible to employ the method of Importance Sampling for approximating to $p(u_i|x_i;\mu_i^*)$ at $x_i = x_i^{(m)}$ for $x_i^{(m)} \in S_{x_i}$ and for all $u_i \in \mathcal{U}_i$ through the instrumental density $p(y_i)$ and utilizing the sample set S_{y_i} together with the importance weights given by

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

and obtain

$$\tilde{p}(u_i|x_i^{(m)};\mu_i^*) = \frac{1}{\sum_{p=1}^{P}\omega_i^{(m)(p)}} \sum_{p=1}^{P} \omega_i^{(m)(p)} \delta_{u_i,\mu_i^*(y_i^{(p)})}$$

for all $u_i \in \mathcal{U}_i$ and for all $x_i^{(m)} \in S_{x_i}$. In other words $\tilde{p}(u_i|x_i^{(m)};\mu_i^*) \approx p(u_i|x_i^{(m)};\mu_i^*)$ and after replacing the latter with the former in Eq.s(49) we achieve $\tilde{P}^*_{i\to j}(u_{i\to j}|x_i^{(m)}) \approx P^*_{i\to j}(u_{i\to j}|x_i^{(m)})$. Similarly, replacing the latter with the former in Eq.(81), we obtain

$$\tilde{P_j^*}^2(u_{ne(j)}|x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M}\omega_j^{(m)(m')}} \sum_{m'=1}^{M} \omega_j^{(m)(m')} \prod_{i\in ne(j)} \tilde{P}^*_{i\to j}(u_{i\to j}|x_i^{(m')}) \tag{82}$$

and after replacing $\tilde{P_j^*}^2(u_{ne(j)}|x_j^{(m)})$ with $P_j^*(u_{ne(j)}|x_j^{(m)})$ in Eq.(80) a futher approximation to the estimation rule is achieved. Let us denote the 3–step approximation to the estimation rule by $\tilde{\nu_j^*}^3(y_j, u_{ne(j)})$, then $\tilde{\nu_j^*}^3(y_j, u_{ne(j)}) \approx \tilde{\nu_j^*}^2(y_j, u_{ne(j)}) \approx \tilde{\nu_j^*}^1(y_j, u_{ne(j)}) \approx \nu_j^*(y_j, u_{ne(j)})$ holds.

Next, we consider evaluating the remaining node-to-node term at the required points of its domain utilizing Eq.s(52) and (53). Consider $C^*_{i\to j}$ and suppose that for $i \in ne(j)$ $I_i^*(u_{ne(i)}, x_i;\nu_i^*)$ is known for all $u_{ne(i)} \in \mathcal{U}_{ne(i)}$ and $x_i \in \mathcal{X}_i$ (Note that, in Proposition 3.7, $\gamma_{\setminus j}$ is fixed at $\gamma_{\setminus j} = \gamma^*_{\setminus j}$). We also assume that for all $j' \in ne(i) \setminus j$, $P^*_{j'\to i}(u_{j'\to i}|x_{j'})$ is known for all $u_{j'\to i} \in \mathcal{U}_{j'\to i}$ and $x_{j'} \in \mathcal{X}_{j'}$. However, the right hand side of Eq.s(52) still does not yield a solution that can be practically carried out in general and we resort to Monte Carlo methods in order to approximately evaluate $C^*_{i\to j}$ at the required points of its domain.

Let us construct a new sample set by concatenating the m^{th} samples of the usual sample sets of neighbors of i other than j, i.e., $S_{x_{j'}}$ for $j' \in ne(i) \setminus j$ given by

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

We apply the same procedure with S_{x_i} and $S_{x_{ne(i)\setminus j}}$ yielding $S_{x_{i\cup ne(i)\setminus j}} = \{x^{(m)}_{i\cup ne(i)\setminus j}\}$ and observe that $x^{(m)}_{i\cup ne(i)\setminus j} \sim p(x_i)\prod_{j'\in ne(i)\setminus j} p(x_{j'})$ for all $x^{(m)}_{i\cup ne(i)\setminus j} \in S_{x_{i\cup ne(i)\setminus j}}$. Then, it is possible to utilize this sample set for an Importance Sampling approximation implying the importance density $p(x_i)\prod_{j'\in ne(i)\setminus j} p(x_{j'})$ together with the importance weights

$$\omega_i^{(m)(m')} = \frac{p(x^{(m')}_{ne(i)\setminus j}, x_i^{(m')} | x_j^{(m)})}{p(x_i^{(m')}) \prod\limits_{j'\in ne(i)\setminus j} p(x_{j'}^{(m')})}$$

and obtain

$$\tilde{C}^*_{i\to j}(u_{j\to i}, x_j^{(m)}) = \sum_{u_{ne(i)\setminus j}} \frac{1}{\sum_{m'=1}^M \omega_i^{(m)(m')}} \sum_{m'=1}^M \omega_i^{(m)(m')} \times \prod_{j'\in ne(i)\setminus j} P^*_{j'\to i}(u_{j'\to i} | x_{j'}^{(m')}) I_i^*(u_{ne(i)}, x_i^{(m')}; \nu_i^*) \quad (83)$$

After replacing $C^*_{i\to j}$ with $\tilde{C}^*_{i\to j}$ in the one–step approximated local communication rule local to node j given by Eq.(79), we obtain $\tilde{\mu}_j^{*\,2}$ such that $\tilde{\mu}_j^{*\,2}(y_j) \approx \tilde{\mu}_j^{*\,1}(y_j)$ for all $y_j \in \mathcal{Y}_j$ with non-zero probability.

Having proposed approximations for the node–to–node terms, we finally handle the evaluation of $I_i^*(u_{ne(i)}, x_i; \nu_i^*)$ at all $u_{ne(i)} \in \mathcal{U}_{ne(i)}$ and $x_i = x_i^{(m)}$ for all $x_i^{(m)} \in S_{x_i}$ that is required in Eq.(83). Note that substituting Eq.(8) in Eq.(53) yields

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

for which the utilization of the sample set S_{y_i} implies an Importance Sampling approximation using the instrumental density $p(y_i)$ together with the importance weights

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

given by

$$\tilde{I}_i^*(u_{ne(i)}, x_i^{(m)}; \nu_i^*) = \frac{1}{\sum\limits_{p=1}^P \omega_i^{(m)(p)}} \sum_{p=1}^P \omega_i^{(m)(p)} c_i^d(\nu_i^*(y_i^{(p)}, u_{ne(i)}), x_i^{(m)})$$

for all $u_{ne(i)} \in \mathcal{U}_{ne(i)}$ and $x_i^{(m)} \in S_{x_i}$ such that $\tilde{I}_i^*(u_{ne(i)}, x_i^{(m)}; \nu_i^*) \approx I_i^*(u_{ne(i)}, x_i^{(m)}; \nu_i^*)$. Replacing I_i^* with $\tilde{I}_i^*$ in Eq.(83) and Eq.(79), we obtain $\tilde{\mu}_j^{*\,3}$ such that $\tilde{\mu}_j^{*\,3}(y_j) \approx \tilde{\mu}_j^{*\,2}(y_j) \approx \tilde{\mu}_j^{*\,1}(y_j)$ for all $y_j \in \mathcal{Y}_j$ with non-zero probability.

D. MC optimization of two–stage in–network processing strategies over UGs

In Sections V-A–V-C,similar to that presented in Section III-A for in–network processing strategies constrained by DAGs, we have provided a Monte Carlo framework for approximating the j^{th} person–by–person optimal local

rule when all the other rules are fixed at the optimal, i.e., $\gamma_{\setminus j} = \gamma^*_{\setminus j}$. for decentralized estimation networks constrained by DAGs. In particular, regarding Proposition 3.7 and given $\gamma^*_{\setminus j} \in \Gamma^{\mathcal{G}}_{\setminus j}$, the proposed framework yields $\tilde{\gamma}^*_j = (\tilde{\mu^*_j}^3(y_j), \tilde{\nu^*_j}^3(y_j, u_{ne(j)}))$ such that $\tilde{\mu^*_j}^3(y_j) \approx \mu^*_j(y_j)$ and $\tilde{\nu^*_j}^3(y_j, u_{ne(j)}) \approx \nu^*_j(y_j, u_{ne(j)})$ for all $y_j \in \mathcal{Y}_j$ and $u_{ne(j)} \in \mathcal{U}_{ne(j)}$ with nonzero probability.

It is possible to utilize the approximations for all local rules, i.e., γ_j for all $j \in \mathcal{V}$, and the node–to–node terms would require the usual sample sets utilized for one–step approximations to the local rules. In addition, the particle representations and approximate computations are valid for any set of two-stage local rules over an undirected graph, including those in an "approximating" form. Let us summarize the Monte Carlo framework with

$$\begin{aligned}
\tilde{\alpha}_j(S_{x_j}) &= \tilde{r}_j(\tilde{C}_{ne(j)\to j}) \\
\tilde{\beta}_j(S_{x_j}, \hat{x}_j) &= \tilde{q}_j(\tilde{P}_j(S_{x_j})) \\
\tilde{P}_j(S_{x_j}) &= \tilde{f}_j(\tilde{P}_{ne(j)\to j}) \\
\tilde{P}_{j\to ne(j)} &= \tilde{g}_j(\tilde{\alpha}_j(S_{x_j}) \\
\tilde{C}_{j\to ne(j)} &= \tilde{h}_j(\tilde{\beta}_j(S_{x_j}, \hat{x}_j), \tilde{P}_{ne(j)\to j})
\end{aligned}$$

where

$$\tilde{\alpha}_j(S_{x_j}) = \{(\lambda c^c_j(u_j, x_j) + \sum_{i\in ne(j)} \tilde{C}_{i\to j}(u_{j\to i}, x_j), u_j, x_j) | u_j \in \mathcal{U}_j, x_j \in S_{x_j}\}$$

$$\tilde{C}_{ne(j)\to j} = \{\tilde{C}_{i\to j}(S_{x_j}) | i \in ne(j)\}$$

$$\tilde{C}_{i\to j}(S_{x_j}) = \{(\tilde{C}_{i\to j}(u_{j\to i}, x_j), u_{j\to i}, x_j) | u_{j\to i} \in \mathcal{U}_{j\to i}, x_j \in S_{x_j}\}$$

$$\tilde{P}_j(S_{x_j}) = \{(\tilde{P}_j(u_{ne(j)}, x_j, u_{ne(j)}, x_j) | u_{ne(j)} \in \mathcal{U}_{ne(j)}, x_j \in S_{x_j}\}$$

$$\tilde{P}_{ne(j)\to j} = \{\tilde{P}_{i\to j}(S_{x_i}) | i \in j\}$$

$$\tilde{P}_{i\to j}(S_{x_i}) = \{(P_{i\to j}(u_{i\to j}, x_i), u_{i\to j}, x_i) | u_{i\to j} \in \mathcal{U}_{i\to j}, x_i \in \mathcal{X}_i\}$$

$$\tilde{P}_{j\to ne(j)} = \{\tilde{P}_{j\to i}(S_{x_j}) | i \in ne(j)\}$$

$$\tilde{C}_{j\to ne(j)} = \{\tilde{C}_{j\to i}(S_{x_i}) | i \in ne(j)\}$$

The Monte Carlo optimization scheme which is obtained through employing the framework in the `Update` step of Algorithm 3 is given by Algorithm 5. Finally, the objective value corresponding a strategy $\gamma \in \Gamma^{\mathcal{G}}$, i.e., $J(\gamma) = J_d(\gamma) + \lambda J_c(\gamma)$ given by Eq.s(59)–(61), can be computed approximately by

$$\tilde{J}(\tilde{\gamma}^l) = \sum_{i\in\mathcal{V}} \tilde{G}^d_i(\tilde{\nu}^l_i) + \lambda \sum_{i\in\mathcal{V}} \tilde{G}^c_i(\tilde{\mu}^l_i) \tag{84}$$

where

$$\tilde{G}^d_i(\tilde{\nu}^l_i) = \sum_{u_{ne(i)}, m} \tilde{P}^{l+1}_i(u_{ne(i)} | x^{(m)}_i) \tilde{I}^l_i(u_{ne(i)}, x^{(m)}_i; \tilde{\nu}^l_i) \tag{85}$$

Algorithm 5 Iterations converging to an approximate pbp optimal two–stage in-network processing strategy over an UG $\mathcal{G}$.

1: Choose $\gamma^0 = (\gamma_1^0, \gamma_2^0, ..., \gamma_N^0)$ such that $\gamma_j^0 \in \Gamma_j^{\mathcal{G}}$ for $j = 1, 2, ..., N$; Choose $\varepsilon \in \mathbb{R}^+$;$l = 0$ ▷ `Initialize`

2: $l = l + 1$

3: `For` $j = 1, 2, ..., N$ `Do` ▷ `Update Step 1: Compute message likelihoods`

$$\tilde{P}^l_{j\to ne(j)} = \tilde{g}_j(\tilde{\alpha}_j^{l-1})$$

4: `For` $j = 1, 2, ..., N$ `Do` ▷ `Update Step 2: Update the stage--two rules`

$$\tilde{P}^l_j(S_{x_j}) = f_j(\tilde{P}^l_{ne(j)\to j})$$

$$\tilde{\beta}^l_j = \tilde{q}_j(P^l_j)$$

$$\tilde{C}^l_{j\to ne(j)} = \tilde{h}_j(\tilde{\beta}_j, \tilde{P}^l_{ne(j)\to j})$$

5: `For` $j = 1, 2, ..., N$ `Do` ▷ `Update Step 3: Update the stage--one rules.`

$$\tilde{\alpha}^l_j = \tilde{r}^l_j(\tilde{C}_{ne(j)\to j})$$

6: `If` $\tau(\tilde{J}(\tilde{\gamma}^l), \tilde{J}(\tilde{\gamma}^{l-1}), ..., \tilde{J}(\tilde{\gamma}^0)) < \varepsilon$ `STOP, else GO TO` 2 ▷ `Check`

and

$$\tilde{G}^c_i(\tilde{\mu}^l_i) = \sum_{u_i,m} c^c_i(u_i, x_i^{(m)}) p(u_i|x_i^{(m)}; \tilde{\mu}^l_i) \tag{86}$$

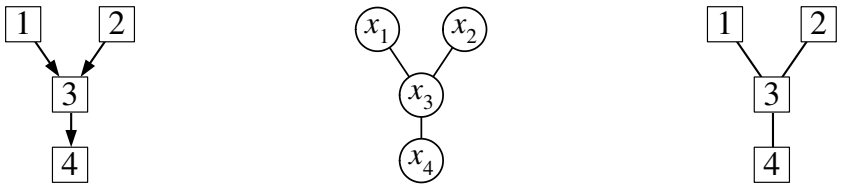
Similar to the discussion presented in Section IV-D for the DAG case, in contrary to $\{J(\gamma^l)\}$, the sequence of approximated objectives, i.e., $\{\tilde{J}(\tilde{\gamma}^l)\}$, is not necessarily non–increasing and considering the error sequence $err[l] = J(\gamma^l) - \tilde{J}(\tilde{\gamma}^l)$ will be identically zero with probability one as $M, P \to \infty$. Investigation of an operator τ (`Check` step of Algorithm 5) that would yiled a non–incereasing error sequence with high probability for finite M, P is beyond the scope of this work.

VI. Examples

In this section, we demonstrate the Monte Carlo optimization algorithms, i.e., Algorithms 4 and 5, introduced in Section IV and V respectively, in various scenarios including Gaussian priors, non-Gaussian priors, and large random graphs.

A. A Simple Gaussian Example

We consider a small network example in which a decentralized estimation network composed of four platforms perform an estimation task. A Gaussian random field $X = \{X_1, X_2, X_3, X_4\}$ is of concern and platform j is associated with X_j. In the first scenario, we consider the underlying communication structure represented by the polytree in Fig. 3(a) , a structure not covered by the star–topology paradigms (e.g., [13] and [19]), as well as stringent



(a) (b) (c)

Fig. 3. (a) Illustration of the DAG $\mathcal{G} = (\mathcal{V}, \mathcal{X})$ where $\mathcal{V} = \{1, 2, 3, 4\}$ and $\mathcal{E} = \{(1, 3), (2, 3), (3, 4)\}$, (b) illustration of the Markov Random Field $\mathcal{G}_X$ subject to estimation by the decentralized estimation network, (c) illustration of the UG that the decentralized estimation strategy is based in the example scenario.

BW constraints such that $\mathcal{U}_{1\to3} = \mathcal{U}_{2\to3} = \mathcal{U}_{3\to4} = \{0, 1, 2\}$. We call this a 1–bit selective communication scheme and also consider 2–bit and 3–bit schemes to be discussed later in this section. We call this a 1–bit selective communication scheme and also consider 2–bit and 3–bit schemes to be discussed later in this section. The online processing scheme operates as given in Section II-A: Since nodes 1 and 2 are parentless, upon measuring y_1 and $y_2 \in \mathbb{R}$ induced by X_1 and X_2, they evaluate their local rules as $(u_{1\to3}, \hat{x}_1) = \gamma_1(y_1)$ and $(u_{2\to3}, \hat{x}_2) = \gamma_2(y_2)$ respectively. Upon receiving these messages and measuring $y_3 \in \mathbb{R}$ induced by X_3, node 3 evaluates its local rule $(u_{3\to4}, \hat{x}_3) = \gamma_3(y_3, u_{1\to3}, u_{2\to3})$, and similarly node 4 evaluates $\hat{x}_4 = \gamma_4(y_4, u_{3\to4})$. The strategy $\gamma = (\gamma_1, ..., \gamma_4)$ is subject to design, which we perform through Algorithm 4.

In addition we comply with Assumption 3 and select separable local costs also enabling Assumption 5 to hold. The cost function local to node j is given by $c_j(u_j, \hat{x}_j, x_j) = c_j^d(x_j, \hat{x}_j) + \lambda c_j^c(u_j, x_j)$ and

$$c_j^c(u_j, x_j) = \sum_{k\in\chi(j)} c_{j\to k}^c(u_{j\to k}, x_j)$$

where $c_{j\to k}^c(u_{j\to k})$ is the cost of transmitting the symbol $u_{j\to k}$ on the link $(j, k) \in \mathcal{E}$. It is selected as

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

indicating the link use. Hence, $\mathcal{U}_{j\to k}$ together with $c_{j\to k}^c$ define 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. The estimation error is penalized by $c_j^d(x_j, \hat{x}_j) = (x_j - \hat{x}_j)^2$. 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 of concern is a multivariate Gaussian, i.e., $x \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_X)$, and Markov with respect to the graph $\mathcal{G}_X$ presented in Fig. 3(b). The covariance matrix is given by

$$\mathbf{C}_X = \begin{bmatrix} 2 & 1.125 & 1.5 & 1.125 \\ 1.125 & 2 & 1.5 & 1.125 \\ 1.5 & 1.5 & 2 & 1.5 \\ 1.125 & 1.125 & 1.5 & 2 \end{bmatrix} \tag{87}$$

which conforms with the Markov properties of $\mathcal{G}_X$. Altough the communication structure of the decentralized estimation network is not related with the MRF representation of X and Algorithm 4 would produce results for any choice, for the sake of simplicity we selected the graph in Fig. 3(b) as the undirected counterpart of that in Fig. 3(a).

The noise processes n_j for $j \in \mathcal{V}$ are additive, mutually independent and given by $n_j \sim \mathcal{N}(0, 0.5)$, so that Assumption 1 holds. In addition, we suppose that Assumption 2 holds and the observation likelihoods are $p(y_j|x_j) = \mathcal{N}(x_j, 0.5)$. Considering $\mathbf{C}_X$, each sensor has an SNR of 6dB.

Since separable local cost functions are utilized, the pbp optimal rules are also split into estimation and communications functions given by Eq.(31) and (32) respectively. Moreover, owing to the squared error local estimation penalty given by c_j^d, the local estimation rules take the form given in Eq.(34). We initialize the local rules, i.e., ν_j^0 and μ_j^0 for $j \in \mathcal{V}$, as follows:

1) Each node applies a myopic inference rule, i.e., performs estimation solely based on its local measurements. This rule is selected as the MMSE estimation rule, i.e., $E\{X_j|Y_j = y_j\}$ given by

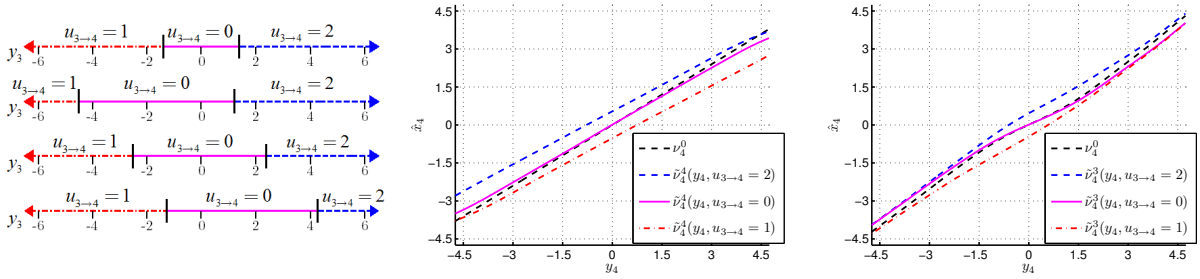
$$\nu_j^0(y_j, u_{\pi(j)}) = \int_{-\infty}^{\infty} \mathrm{d}x_j \; x_j p(x_j|y_j) \tag{88}$$

2) All the nodes apply an initial communication rule as a quantization of the observation y_j, i.e.,

$$\mu_i^0(y_i, u_{\pi(i)}) = \begin{cases} 1 \ , \ y_i < -2\sigma_n \\ 0 \ , \ -2\sigma_n \leqslant y_i \leqslant 2\sigma_n \\ 2 \ , \ y_i > 2\sigma_n \end{cases} \tag{89}$$

Considering $J(\gamma) = J_d(\gamma) + \lambda J_c(\gamma)$ and pbp optimal strategies achieved through Algorithm 2, in principle, different values of λ would yield different performance points $(J_c(\gamma^*), J_d(\gamma^*))$. Moreover, in this case, after a certain value $\lambda = \lambda^*$, the communication cost λJ_c will dominate such that the decrease in the decision cost J_d with the contributions of the communicated symbols will not be enough to decrease J and symbol 0 will be the best choice. Moreover, the individual estimators will be the myopic rules, since myopic rules with no communications constitute a 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 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.

We use 2000 and 30000 samples for each S_{x_i} and S_{y_i} generated from $p(x_i)$ and $p(y_i)$ respectively and use Algorithm 4 for varying λ from 0 with 0.001 steps. Example converged local communication and estimation rules are presented in Fig. 4(a) and (b) for node 3 and 4 respectively, where $\lambda = 0.1$ and convergence is declared after 4 "offline" iterations. Note that the initial communication rule shown at the top row of Fig. 4(a) and the initial estimation rule illustrated by the black dashed line in Fig. 4(b) are valid for all of the nodes with appropriate choices of the domain and range labels. The pbp optimal communication rule local to node 3 can be treated as a collection of threshold rules for each incoming message value (some of which are illustrated in Fig. 4(a)). Now, let us turn



(a) (b) (c)

Fig. 4. Example converged rules regarding the Gaussian DAG problem ((a),(b)) and the Laplacian DAG problam ((c)). (a) Communication rules for node 3: (from top to bottom) the initial communication rule, i.e., $u_{3\to4}=\mu_3^0(y_3,u_{\pi(3)})$ and illustrations of the converged communication rule for the Gaussian example for $\lambda=0.1$ at the end of the 4^{th} step, specifically, $u_{3\to4}=\mu_3^4(y_3,u_{\pi(3)})$ for $u_{\pi(3)}=\{2,2\},\{0,0\}$ and $\{1,1\}$ respectively. (b) Illustrations of the initial and converged estimation rules for node 4 for the Gaussian example at the end of the 4^{th} step, i.e., ν_4^0 and $\hat{x}_4=\tilde{\nu}_4^4(y_4,u_{\pi(4)})$ respectively. (c) For the Laplacian example; the converged estimation rule local to node 4 at the end of the 3^{rd} step, i.e., $\hat{x}_4=\tilde{\nu}_4^3(y_4,u_{\pi(4)})$.

to the estimation rule in Fig. 4(b). If the message received by node 4 suggests a high/medium/low value for x_4 that is consistent with y_4, then the pbp estimation rule local to node 4 acts similar to the myopic rule (Note the asymptotic behaviour of $\tilde{v}_4^4$ for $u_{3\to4}=2$ and $u_{3\to4}=1$ respectively in comparison with the initial rule as well as $\tilde{v}_4^4$ for $u_{3\to4}=0$ in Fig. 4(b)), otherwise, the estimate diverts from the nominal values as implied by the incoming message.

The approximate performance points, i.e., $(\tilde{J}_c,\tilde{J}_d)$ pairs where $\tilde{J}_c$ is the approximate total link use rate and $\tilde{J}_d$ is the approximate total MSE, of the converged strategies $\tilde{\gamma}^*$ are presented in Fig. 3(a)(black '+'s). The upper and lower bounds are MSEs corresponding to the myopic rule and the centralized optimal rule respectively. We repeat the same scenario with different BW constraints: Specifically, we select $\mathcal{U}_{i\to j}$s corresponding to 2 and 3–bit selective communication schemes. The initial communication rules are appropriately modified versions of that given by Eq.(89) and the approximate performance points obtained are presented in Fig. 3(a) as well. We use the condition $\left|\tilde{J}(\tilde{\gamma}^{l-1})-\tilde{J}(\tilde{\gamma}^l)\right|<1.0e-4$ in the `Check` step of Alg.4. The average number of steps for convergence (within $\pm3\sigma$) are 3.6 ± 1.5, 4.2 ± 2.0 and 4.1 ± 1.8 for $1,2$ and 3–bit schemes respectively. Note that, 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. Considering $(\tilde{J}_c,\tilde{J}_d)$ pairs for the 1–bit selective communication scheme, for $\lambda=0$, the transmission has no cost, but the link use rate is well below 75% of the total 3 bits. This indicates that the information of receiving no messages is successfully maintained in this perspective. Moreover, the communication stops for $\lambda^*\approx0.355$. Similarly, approximate points for 2–bit and 3–bit schemes indicate that, if λ is small enough, we can achieve smaller MSE for the same total communication load as we increase the link capacities.

Next, we consider a two–stage strategy over the undirected graph given in Fig. 3(c) for the same estimation problem. The set of admissible symbols is given by $\mathcal{U}_{i\to j}=\{0,1,2\}$ for all $(i,j)\in\mathcal{E}$. In contrast with the directed

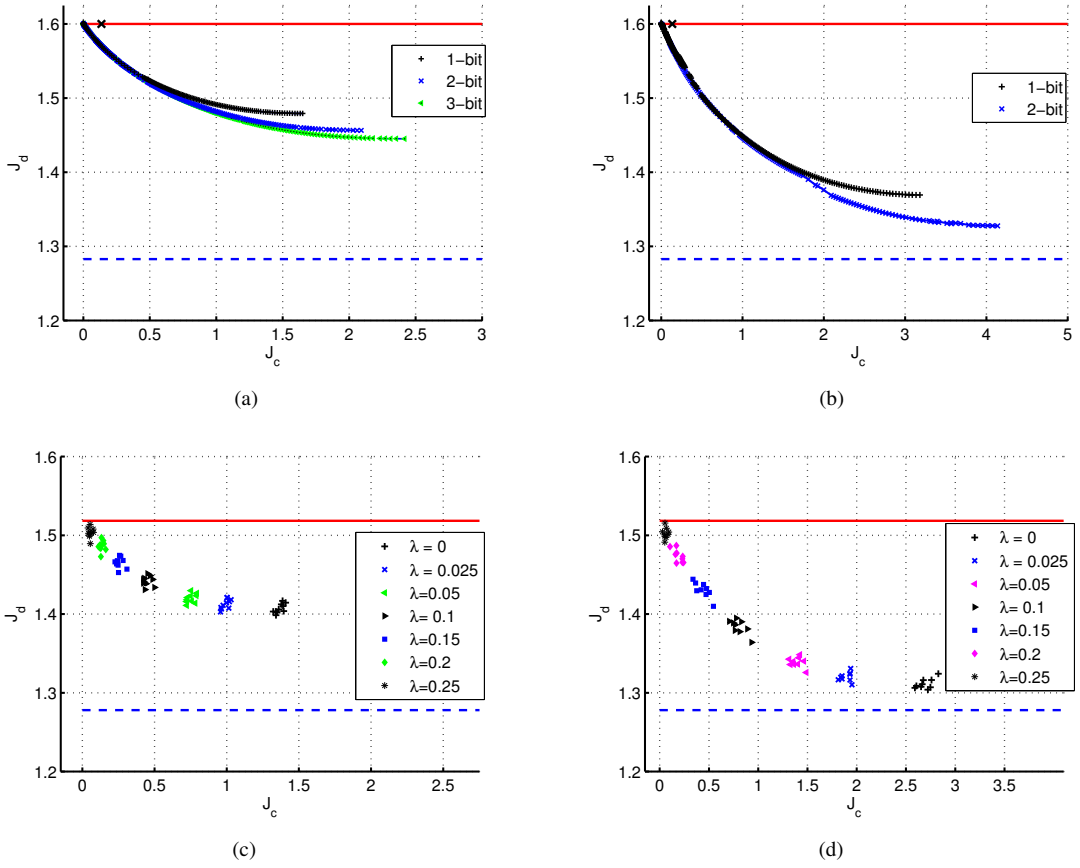


Fig. 5. The approximate performance points converged revealing the tradeoff together with the lower bounds (blue dashed-lines) and the upper bounds (red dashed-lines) of the problems given by the estimation performance measured in MSE for the optimum centralized and the myopic rules respectively.

(a) Gaussian DAG problem: The estimation network in Fig. 3(a) is subject to optimization through Alg. 4 starting with the initial rules given by Eq.s(88) and (89) which achieve $(J_c(\gamma^0), J_d(\gamma^0))$ (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 $1, 2$ and 3 bit selective communication schemes are presented.

(b) Gaussian UG problem: The estimation network in Fig. 3(c) is subject to optimization through Algorithm 5 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.

(c) Heavy tailed (Laplacian) prior problem with a DAG: Approximate performance points are presented which are achieved for the heavy tailed prior case, for various values of λ and 10 sample sets for each λ through Alg. 4.

(d) Heavy tailed (Laplacian) prior problem with a UG: Approximate performance points are presented which are achieved for the heavy tailed prior case, for various values of λ and 10 sample sets for each λ through Alg. 5.

case, the online processing starts with each node evaluating its stage–one communication function on its measurement, i.e., $u_{1\to3} = \mu_1(y_1)$, $u_{2\to3} = \mu_2(y_2)$, $(u_{3\to1}, u_{3\to2}, u_{3\to4}) = \mu_3(y_3)$ and $u_{4\to3} = \mu_4(y_4)$ simultaneously. As soon as all the messages from the neigbors (or lack thereof) are received, stage two estimation rules are evaluated as $\hat{x}_1 = \nu_1(y_1, u_{3\to1})$,$\hat{x}_2 = \nu_2(y_2, u_{2\to3})$, $\hat{x}_3 = \nu_3(y_3, u_{1\to3}, u_{2\to3}, u_{4\to3})$ and $\hat{x}_4 = \nu_4(y_4, u_{3\to4})$. We design the

strategy $\gamma = (\gamma_1, ..., \gamma_4)$ where $\gamma_j = (\mu_j, \nu_j)$ using Algorithm 5.

The cost functions are those utilized for the DAG case and similarly, 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 similar to that used in the DAG case with the difference that the j^{th} rule takes as argument the messages from all of the neighbors of node j.

Similarly we approximate to the samples from the pareto–optimal performance curve as we increase λ from 0 and obtain a strategy using Algorithm 5, which in turn provides a quantification for the trade–off between the cost of estimation errors and communication.

In Fig. 5(b) 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 bounds are mean squared errors (MSEs) corresponding to the myopic rule and the centralized optimal rule[10]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 capacity of 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.

Comparing the approximated performance points of the directed and undirected strategies presented in Fig. 5(a) and (b) respectively for 1 bit and 2 bits selective communication schemes, we observe the benefits of bi–directional communications employed by the strategy over the undirected graph. For the directed case, nodes 1 and 2 are parentless and hence do not have means to exploit contributions from other platforms. Specifically all parentless nodes apply the initial rule, which has been selected as the myopic estimator. Therefore the nodes with more ancestors are more likely to benefit the contribution of other nodes whereas for the undirected case, the nodes with more neighbors pose advantegous. The price paid is that the information horizon is limited with the observation of the neighbors whereas the local rules depend on a two–hop neighborhood due to the two stage mechanism necessary for causality.

B. A Simple Heavy Tailed Example

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

$$p(x) = \frac{2}{(2\pi)^{d/2}|C_x|^{1/2}} \left(\frac{x^T C_x^{-1} x}{2}\right)^{1-d/2} K_{1-d/2}(\sqrt{2x^T C_x^{-1} x}) \tag{90}$$

[10]For $c(x, \hat{x}) = (x - \hat{x})^T (x - \hat{x})$, the optimal centralized estimate is the mean vector 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 represent y_j before transmission.

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. [48]). Let us denote this distribution 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 imply a Markov random field.

On the other hand, it is possible to generate samples from a multivariate symmetric Laplacian 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. Given $u \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_X)$ and $z \sim e^{-z}$, generate samples of X by $x = \sqrt{z}u$, then $x \sim SL_d(\mathbf{C_x})$. Therefore, it is possible to express $SL_d(\mathbf{C}_X)$ as

$$p(x) = \int_0^\infty \mathcal{N}(0, z\mathbf{C}_X)p(z)\mathrm{d}z \tag{91}$$

where $p(z) = e^{-z}$. This form, being a scaled sum of Gaussians, generalizes Gaussian mixtures and hence is also referred to as a *scale mixture of Gaussians*[11].

Similar to that in the previous section, we assume the underlying communication structure described by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in Fig. 3(a) together with a 1–bit selective communication scheme on each link, and similar cost functions, observation likelihoods, and initial local rules.

The Monte Carlo framework extends trivially for (finite) Gaussian Mixture Models which can be used to represent arbitrary priors. To the best knowledge of the authors, in the case of an MSL prior, even the centralized paradigm fails to provide a solution without employing numerical approximations.

For our case, 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.(87) and we exploit the fact that the j^{th} marginal distribution of $SL_d(\mathbf{C}_X)$ is given by $SL_1([\mathbf{C}_X]_{j,j})$ and it is straightforward to generate samples from these marginals [51]. For the observations, although the marginal densities yield closed form expressions[12], it is not easy to sample from this density since it does not yield a distribution function in closed form. However, considering the mixture approximation $\sum_{x_j^{(m)} \in S_{x_j}} p(y_j|x_j^{(m)}) \approx p(y_j)$ where $S_{x_j} = \{x_j^{(1)}, x_j^{(2)}, ..., x_j^{(M)}\}$, it is possible to draw samples from $p(y_j)$ approximately by sampling from $p(y_j|x_j^{(m)})$ for $m = 1, 2, ..., M$ until the density *mixes*.

We generate $\left|S_{x_j}\right| = 3000$ samples from the prior marginals and $\left|S_{y_j}\right| = 45000$ samples from the aforementioned mixture densities. We run Algorithm 4 for different choices of λ and for 10 different sample sets. An example converged estimation rule is illustrated in Fig. 4(c) which is local to node 3 and convergenced for $\lambda = 0.1$ after 3 offline iterations. Note that, contrary to that in the Gaussian example, the initial myopic estimation rule for any node is not linear (black dashed curve in Fig. 4(c)) and is successfully represented within the MC framework. The asymptotic behaviours in the case that the measurement and the incoming message confirm each other are similar to that in the Gaussian example.

[11]This family of distributions has been employed to model multiple variables that exhibit uncorrelatedness yet dependence such as the statistics of natural images (see, e.g., [49] and [50]).

[12]It can be shown that $p(y_j) = 0.1410\sqrt{\pi}e^{-y_j+1/4}\left(e^{2y_j} + 1 - \Phi\left(y_j + 1/2\right)e^{2y_j} + \Phi\left(y_j - 1/2\right)\right)$ for $j \in \mathcal{V}$ where Φ is the error function.

In Fig. 5(c), approximate performance points for the converged strategies are presented where the upper and lower bounds are the MSEs corresponding to the myopic and centralized rules respectively. For each λ we observe a cluster around the corresponding point from the pareto–optimal curve with a reasonable variability over sample sets (Fig. 5(c)). This is in accordance with the expectation that heavy tailed distributions require utilization of larger sample sets. Nevertheless, the framework we propose produces distributed solutions in problem settings which do not admit straightforward solutions even in the centralized case.

Next, we present Algorithm 5 in a similar setting. The undirected graph of concern is given in Fig. 3(c). For various values of λ, and the Algorithm is run for 10 different sample sets. The initial rules, costs and the likelihood is similar to those used for the Gaussian UG example in the previous section.

The approximated performance points are presented in Fig. 5 (d). Similar to the Gaussian case, for small values of λ, a decentralized strategy with a comparable performance to the centralized rule is achieved. The benefits of the undirected topology is apparent comparing the MSE performance with that presented in Fig. 5 (c) for the directed case.

C. Examples with Large Graphs

In this section, we demonstrate Algorithms 4 and 5 in relatively large scale problems: 50 platforms are randomly deployed over an area of 100 unit squares and each location $s_j \in \mathbb{R}^2$ is associated with a scalar random variable, x_j. We assume that the random field $X = (X_1, X_2, ..., X_{50})^T$ is Gaussian with zero mean, i.e., $X \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_x)$ and $\mathbf{C}_x = [C_{i,j}]$ complies with the Matern covariance function which is commonly utilized in spatial data modeling [10] and given by

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

where $h \triangleq \|s_i - s_j\|$, K_η is a modified Bessel function of the second kind of order η and τ^2, σ^2 are parameters that determine the decaying characteristics. Such a covariance matrix is presented in Fig. 6(a) corresponding to an example random deployment for which the Gabriel graph is presented in Fig. 6(b). We generate a polytree by randomly selecting 6 childless nodes and employing Kruskal's algorithm on this graph (Fig. 6(c)).

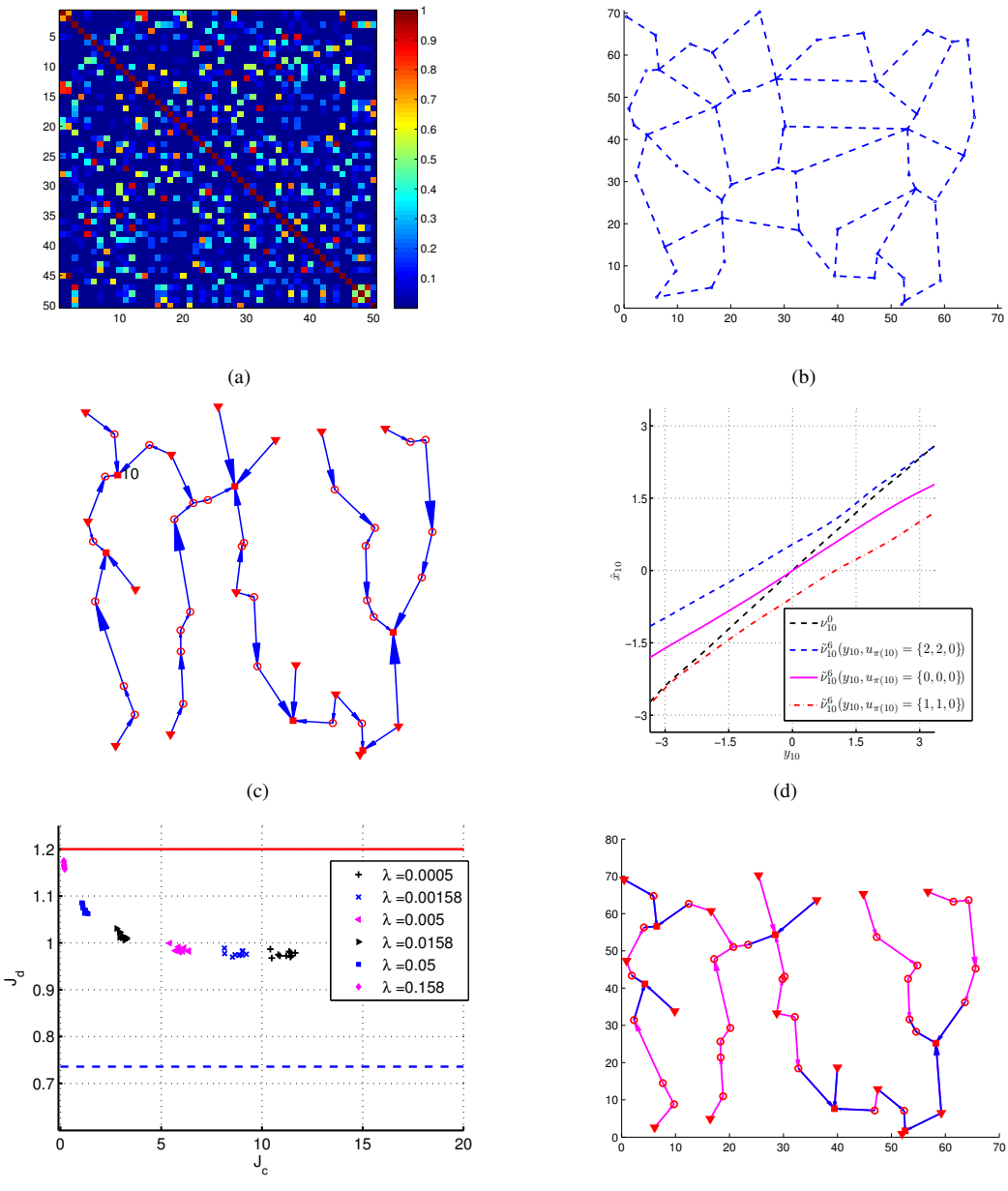


Fig. 6. Regarding the 50 randomly deployed nodes: (a) $\mathbf{C}_x$ obtained through the Matern covariance function. (b) UG generated as a sparsified Gabriel Graph of the deployment. (c) A polytree generated from a spanning tree of the Gabriel Graph of the deployment after randomly selecting 6 childless nodes; parentless and childless nodes are shown by red triangles and red squares (e.g. node 10) respectively. (d) Converged estimation rule local to (childless) node 10 for $\lambda = 0.005$ at the end of 6 iterations. (e) approximate performance points of converged strategies for $\lambda = 0.0005, 0.00158, ..., 0.158$ and 10 sample sets. The upper and lower bounds of the problem are the myopic and the centralized MSEs shown by the solid red line and the dashed blue line respectively. (f) Polytree with links that are not utilised by the approximate pbp optimal strategy obtained after convergence (magenta edges).

Different from the previous scenarios, only the variables associated with the childless nodes are of concern and only the childless nodes perform estimation whereas the remaining operate in a fusion setting such that they only

provide information to the children based on the incoming messages from the parent nodes and the measurement they make. We consider a 1–bit selective communication scheme on each link and the communication cost considers the link use rate. The estimation error is $c_j^d(x_j, \hat{x}_j) = (x_j - \hat{x}_j)^2$ if j is childless and identically zero otherwise. Similar to the previous examples, the initial communication rules are quantization of the observations and the childless nodes are initiated with the corresponding myopic estimation rules (for $\tau^2 + \sigma^2 = 1$ and $\sigma_{n_j}^2 = 0.25$ for all $j \in \mathcal{V}$).

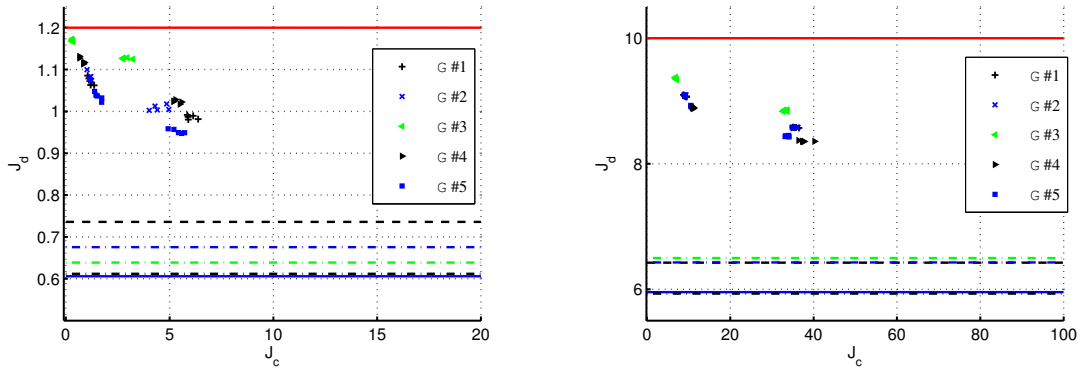
We employ Algorithm 4 for a geometrically increasing sequence $\lambda = 0.0005, 0.00158, ..., 0.158$ and for 10 different sample sets such that $\left|S_{x_j}\right| = 2000$ and $\left|S_{y_j}\right| = 30000$ (over the polytree in Fig. 6(c)). An example converged estimation rule is illustrated in Fig. 6(d). We consider node 10 in Fig. 6(c); the initial myopic rule is linear with the observation y_{10}, however, the converged strategy, as expected considering the previous examples, exhibits a highly nonlinear behaviour as the incoming messages suggest less likely (high or low) values for x_{10}. When no messages is sent, the pbp optimal rule is similar to a mid–way between the estimator functions selected when incoming messages imply a high and a low value for x_{10} respectively.

The Monte Carlo estimates of the performances of approximate pbp optimal strategies are shown in Fig. 6(e). Note that the myopic MSE for each platform is 0.2 yielding a total of 1.2 whereas the centralised MSE (blue dashed–line) is specified by the deployment (through $\mathbf{C}_x$). The MC framework successfully performs in large graph scenarios and makes it possible to identify clusters around points from the pareto–optimal curve capturing the trade–offs.

One observation is that the approximate pbp optimal strategy converged through the MC framework might stop communicating over certain links, even if the approximated strategy does not yield any dead links. This is due to failing to represent, e.g., quantization rules that utilize thresholds exceeding the bounds of the produced samples. For example, suppose the pbp optimal local rule requires thresholds t_1, t_2 such that $t_1 < \min(S_{y_j})$ and $t_2 > \max(S_{y_j})$ for a particular set of incoming messages. Then the approximate local communication rule represents this rule through stoping the communication completely if that particular set of incoming messages is received. An example is presented in Fig. 6(f) in which the converged approximate pbp optimal strategy selects to stop all transmissions over the magenta edges. In other words, over the magenta edges, no transmissions are made for any set of incoming messages.

Nevertheless, the Monte Carlo framework we proposed successfully produces results for ramdon large graphs. Next, we consider 5 different graphs and employ Algorithm 4 for $\lambda = 0.005, 0.05$ and 5 different sample sets for each value. We consider a 1-bit selective communication scheme over each link which yields a total network capacity of 49 bits. Note that for each graph the pareto–optimal curve as well as the lower bound would differ. The Monte Carlo estimates of the performance points of the approximate pbp strategies are given in Fig. 7(a). The mean number of iterations for convergence is 4 (with $\sigma = 0.9$ any heavy on the values greater than 4).

In a similar setting, we consider the two-stage strategies over undirected graphs and assume that the underlying structure is a subgraph of the Gabriel graph corresponding to the random deployment (e.g., Fig. 6(b)). The initial communication rules as well as the costs are similar to that for the random large graph experiments considering a



(a) (b)

Fig. 7. The approximate performance points of the strategies converged by (a) Algorithm 4 for 5 polytrees generated for random deployments and for 5 sample sets for each deployment. 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 the childless nodes. (b) Algorithm 5 for 5 UGs for random deployments and for 5 sample sets for each deployment. 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 centralised MSE (showed by a solid red–line in both figures) is same for all deployments whereas the myopic MSE (the lower bound) vary for each deployment.

DAG except that all nodes perform estimation in this setting. Note that for a 1-bit selective communication scheme over each direction over an undirected link, each graph differs in the total network capacity. For the case, the capacities corresponding to UG 1–5 are $132, 128, 130, 134$and 140 bits respectively. The approximate performance points of the strategies obtained through Algorithm 5 for $\lambda = 0.005, 0.05$ are presented in Fig. 7(b). The number of iterations for convergence has a minimum and mean value of 3 and 4 respectively (with $\sigma = 1.1$ heavy over the values greater than 4).

VII. Conclusion

In this work, we have considered the design of decentralized estimation strategies. Motivated by sensor network applications, we take the communication constraints into account including the availability and BW of the links as well as the cost of transmitting symbols over them. We are particularly interested in trading off estimation accuracy with the utilization of communication resources. We employ two classes of *online* processing strategies over graphs: The first class is constituted of local rules operating in accordance with a (forward) message–passing structure on a DAG. For the second class of strategies, the nodes operate in two–stages rendering a UG. These two classes provide a number of benefits compared with the conventional approaches in decentralized estimation including that they cover any association of the nodes with the variables that make up the global state and they are valid for any corresponding graph representation DAG or UG presumably supported by the available set of links. Another important feature is that, under a Bayesian setting, they yield a rigorous design problem and tractable *offline* strategy optimization procedures in a message passing fashion provided that some reasonable assumptions hold. This design setting, different from that in previous work on decentralized estimation, enables us to explicitly

consider the cost of communications, and for a parametric dual–objective Bayesian risk, a pareto–optimal curve is obtained revealing the trade–off through the graceful degradation of estimation accuracy as the communication becomes more costly. It is also possible to model a broader range of constraints on the communication structure to be used during online processing. For example, it is possible to consider extensions of the conventional star–topology since it is a particular polytree structure. In addition, it is possible to model selective communication schemes through an appropriate selection of the communication cost(s).

The graphical model perspective for decentralized estimation in recent work takes the communication constraints into account to a certain extent, nevertheless a general framework which explicitly captures the cost of transmissions especially under stringent constraints similar to those of our concern has not been introduced. The in–network processing strategies we consider have been previously studied for decentralized detection [27] and hence our first contribution is the extension of these results for the estimation problem and a rephrasing of the offline optimization procedures.

However, in contrast with the detection problem, the global state vector takes values from a Euclidean space in our case, and consequently the forward and backward messages in the offline strategy optimization procedures, as well as the pbp optimal online rules require the computation of integral operators which cannot be evaluated exactly, in general.

We overcome this problem through our second contribution which is a Monte Carlo framework for each class of online processing strategies, under which particle representations together with approximate computational schemes are utilized for all expressions involved, including the local rules. Through this approach, we provide a feasible computational scheme while we conserve the appealing features of the original framework which include scalability with the number of platforms as well as the number of variables involved. The proposed algorithm also scales with the sample set sizes and produces results for any set of distributions provided that samples can be generated from the marginals. We have demonstrated these features through several examples including a Gaussian problem, a non-Gaussian prior problem, and a random large graph scenario in Section VI. The MC optimizations produce reasonable sets of local rules, and we observe that the estimation accuracy is traded–off with communication load as we vary their relative emphases on the total cost. Equivalently, the performances achieved approximate the corresponding pareto–optimal curve. In addition, the proposed optimization approachese can also potentially be applied for hybrid in–network processing strategies employing both families [52].

There are a number of issues left beyond the scope of this work. In contrast with the non–approximated case, the iterative offline strategy optimization procedure does not yield a monotonically decreasing sequence. Investigation of a robust stopping condition remains as future work together with the introduction of possible smoothing approaches through kernel methods. The IS estimate of an integral is known to be mildly biased and the investigation of biasedness of the resulting strategies remains open as well as the solution to the problem of selecting the graph structure that yields the best pbp optimal strategy given an a-priori distribution.

APPENDIX

PROOF OF PROPOSITION 3.3

Provided that Assumption 1 holds, the underlying distribution to the Bayesian framework is given by

$$\begin{aligned} p(u,x,\hat{x};\gamma) &= \int_{\mathcal{Y}} \mathrm{d}y \prod_{i=1}^{N} p(u_i,\hat{x}_i|y_i,u_{\pi(i)};\gamma_i) \prod_{k=1}^{N} p(y_k|x)p(x) \\ &= p(x) \prod_{i=1}^{N} p(u_i,\hat{x}_i|x,u_{\pi(i)};\gamma_i) \end{aligned} \tag{92}$$

which further implies that

$$p(u,\hat{x}|x;\gamma) = \prod_{i=1}^{N} p(u_i,\hat{x}_i|x,u_{\pi(i)};\gamma_i) \tag{93}$$

First, consider Eq.(11) and the term

$$\prod_{i\neq j} p(u_i,\hat{x}_i|x,u_{\pi(i)};\gamma_i) = \prod_{i\neq j} \int_{\mathcal{Y}_i} \mathrm{d}y_i \, p(y_i|x)p(u_i,\hat{x}_i|y_i,u_{\pi(i)};\gamma_i^*)$$

The conditional distribution that equals to the product above is obtained by dividing both sides of Eq.(93) by the contribution of the j^{th} rule, i.e.,

$$\begin{aligned} \prod_{i\neq j} p(u_i,\hat{x}_i|x,u_{\pi(i)};\gamma_i) &= \frac{p(u,\hat{x}|x;\gamma)}{p(u_j,\hat{x}_j|x,u_{\pi(j)};\gamma_j)} \\ &= \frac{p(u_{\setminus\pi(j)},\hat{x}|x,u_{\pi(j)};\gamma)p(u_{\pi(j)}|x;\gamma)}{p(u_j,\hat{x}_j|x,u_{\pi(j)};\gamma_j)} \\ &= p(u_{\setminus j\cup\pi(j)},\hat{x}_{\setminus j}|x,u_{\pi(j)},u_j,\hat{x}_j;\gamma)p(u_{\pi(j)}|x;\gamma) \\ &= p(u_{\setminus j},\hat{x}_{\setminus j}|x,u_j;\gamma_{\setminus j}) \end{aligned} \tag{94}$$

for which after applying the chain and Bayes' rule, we have substituted the conditional independence properties $u_{\pi(j)} \perp\!\!\!\perp (u_j,\hat{x}_j)|x;\gamma_{an(j)}$ where $an(j)$ are the set of ancestor nodes of j and $(u_{\setminus j},\hat{x}_{\setminus j}) \perp\!\!\!\perp \hat{x}_j|x,u_j;\gamma_{\setminus j}$ due to the directed acyclic nature in the last step.

Then we follow similar steps with that for the detection case in [27] whereas in our setting, X takes values from a denumarable set $\mathcal{X}$ and we do not utilize a channel model, i.e., we assume that all links are error free.

Consider Eq.s(11) and (18) together with Eq.(94). After substituting the mathematical expression of the cost locality assumption, i.e., Eq.(19), in Eq.(18) we obtain

$$\begin{aligned} &\theta_j^*(u_j,\hat{x}_j,x,u_{\pi(j)}) \\ &= \sum_{i\in\mathcal{V}} \sum_{u_{\setminus\{j\}\cup\pi(j)}} \int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j} \, c(u_i,x_i,\hat{x}_i)p(u_{\setminus j},\hat{x}_{\setminus j},x|u_j;\gamma_{\setminus j}^*) \\ &= p(u_{\pi(j)}|x;\gamma_{\setminus j}^*)p(x)c(u_j,x_j,\hat{x}_j) + \sum_{i\in\mathcal{V}\setminus j} \sum_{u_{\setminus\{j\}\cup\pi(j)}} \int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j} \, c(u_i,x_i,\hat{x}_i)p(u_{\setminus j},\hat{x}_{\setminus j},x|u_j;\gamma_{\setminus j}^*) \end{aligned}$$

and treat the summation over $i \in \mathcal{V}\setminus j$ in three groups: $de(j) \in \mathcal{V}\setminus j$ denoting the decendants of node j, $\pi(j) \in \mathcal{V}\setminus j$ denoting the parent of node j and $an(j)\setminus\pi(j) \in \mathcal{V}\setminus j$ denoting the ancestors of node j that are not its parents. Due

to the directed acyclic nature, these sets are mutually exclusive. Hence

$$\begin{aligned}
&\sum_{i\in\mathcal{V}\setminus j}\sum_{u_{\setminus\{j\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j}\, c(u_i,x_i,\hat{x}_i)p(u_{\setminus j},\hat{x}_{\setminus j},x|u_j;\gamma^*_{\setminus j}) \\
&\quad= \sum_{m\in de(j)}\sum_{u_{\setminus\{j\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j}\, c(u_m,x_m,\hat{x}_m)p(u_{\setminus j},\hat{x}_{\setminus j},x|u_j;\gamma^*_{\setminus j}) \\
&\qquad\quad+\sum_{k\in\pi(j)}\sum_{u_{\setminus\{j\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j}\, c(u_k,x_k,\hat{x}_k)p(u_{\setminus j},\hat{x}_{\setminus j},x|u_j;\gamma^*_{\setminus j}) \\
&\qquad\qquad+\sum_{n\in an(j)\setminus\pi(j)}\sum_{u_{\setminus\{j\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j}\, c(u_n,x_n,\hat{x}_n)p(u_{\setminus j},\hat{x}_{\setminus j},x|u_j;\gamma^*_{\setminus j})
\end{aligned}$$

Consider the first group on the right hand side of the equation above. The following holds

$$\begin{aligned}
&\sum_{m\in de(j)}\sum_{u_{\setminus\{j\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j}\, c(u_m,x_m,\hat{x}_m)p(u_{\setminus\{j\}\cup\pi(j)},\hat{x}_{\setminus j}|x,u_j,u_{\pi(j)};\gamma^*_{\setminus j})p(u_{\pi(j)}|x;\gamma^*_{\setminus j})p(x) \\
&= \sum_{m\in de(j)}\sum_{u_m}\int_{\mathcal{X}_m} \mathrm{d}\hat{x}_m\, c(u_m,x_m,\hat{x}_m)\sum_{u_{\setminus\{j,m\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus\{j,m\}}} \mathrm{d}\hat{x}_{\setminus\{j,m\}}\, p(u_{\setminus\{j\}\cup\pi(j)},\hat{x}_{\setminus j}|x,u_j,u_{\pi(j)};\gamma^*_{\setminus j})p(u_{\pi(j)}|x;\gamma^*_{\setminus j})p(x) \\
&= p(u_{\pi(j)}|x;\gamma^*_{\setminus j})p(x)\sum_{m\in de(j)}\sum_{u_m}\int_{\mathcal{X}_m} \mathrm{d}\hat{x}_m\, c(u_m,x_m,\hat{x}_m)p(u_m,\hat{x}_m|x,u_j,u_{\pi(j)};\gamma^*_{\setminus j}) \\
&= p(u_{\pi(j)}|x;\gamma^*_{\setminus j})p(x)\sum_{m\in de(j)}\sum_{u_m}\int_{\mathcal{X}_m} \mathrm{d}\hat{x}_m\, c(u_m,x_m,\hat{x}_m)p(u_m,\hat{x}_m|x,u_j;\gamma^*_{\setminus j})
\end{aligned}$$

where we apply the chain rule and rearrange the order of operators except that in the last step we assert the assumption that $\mathcal{G}$ is a polytree when $p(u_m,\hat{x}_m|x,u_j,u_{\pi(j)};\gamma^*_{\setminus j})$ is reduced to $p(u_m,\hat{x}_m|x,u_j;\gamma^*_{\setminus j})$ for $m\in de(j)$. Since the polytree topology implies that there are no paths from any of the ancestors of node j to any of its decendats that does not pass through j, given u_j and having $\gamma^*_{\setminus j}$ determined, $u_{\pi(j)}$ has no bearing on $(u_m,\hat{x}_m)$ where $m\in de(j)$ which would not necessarily be the case if $\mathcal{G}$ were not a polytree.

Considering the summation over the second group, similar rearrangements are performed yielding

$$\begin{aligned}
&\sum_{k\in\pi(j)}\sum_{u_{\setminus\{j\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j}\, c(u_k,x_k,\hat{x}_k)p(u_{\setminus j},\hat{x}_{\setminus j},x|u_j;\gamma^*_{\setminus j}) \\
&\quad= \sum_{k\in\pi(j)}\int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k\, c(u_k,x_k,\hat{x}_k)p(x)\sum_{u_{\setminus\{j\}\cup\pi(j)}}\int_{\mathcal{X}_{\setminus\{j,k\}}} \mathrm{d}\hat{x}_{\setminus\{j,k\}}\, p(u_{\setminus j},\hat{x}_{\setminus j}|x,u_j;\gamma^*_{\setminus j}) \\
&\quad= \sum_{k\in\pi(j)}\int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k\, c(u_k,x_k,\hat{x}_k)p(x)p(u_{\pi(j)},\hat{x}_k|x,u_j;\gamma^*_{\setminus j}) \\
&\quad= \sum_{k\in\pi(j)}\int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k\, c(u_k,x_k,\hat{x}_k)p(x)p(u_k,\hat{x}_k|x,u_j,u_{\pi(j)\setminus k};\gamma^*_{\setminus j})p(u_{\pi(j)\setminus k}|x,u_j;\gamma^*_{\setminus j}) \\
&\quad= \sum_{k\in\pi(j)}\int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k\, c(u_k,x_k,\hat{x}_k)p(x)p(u_k,\hat{x}_k|x;\gamma^*_{\setminus j})p(u_{\pi(j)\setminus k}|x;\gamma^*_{\setminus j}) \\
&\quad= p(x)p(u_{\pi(j)\setminus k}|x;\gamma^*_{\setminus j})\sum_{k\in\pi(j)}\int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k\, c(u_k,x_k,\hat{x}_k)p(u_k,\hat{x}_k|x;\gamma^*_{\setminus j})
\end{aligned}$$

where in the first two steps, we rearrange operators and perform marginalization, in the third step we apply the chain rule. In the fourth step, the u_j and $u_{\pi(j)\setminus k}$ arguments of the conditional drops since due to the polytree topology

no two parents of node j shares a common ascendant and these arguments are non-informative for $(u_k, \hat{x}_k)$ when $\gamma_k^* \in \gamma_{\setminus j}^*$ is determined. Also note that, at the last step, the terms contain no contribution of $(u_j, \hat{x}_j)$ and hence have no bearing on the optimization regarding the person-by-person optimal rule of node j.

A similar treatment of the third group yields

$$\begin{aligned}
&\sum_{n\in an(j)\setminus\pi(j)} \sum_{u_{\setminus\{j\}\cup\pi(j)}} \int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j}\, c(u_n, x_n, \hat{x}_n) p(u_{\setminus j}, \hat{x}_{\setminus j}, x | u_j; \gamma_{\setminus j}^*) \\
&= \sum_{n\in an(j)\setminus\pi(j)} \sum_{u_n} \int_{\mathcal{X}_n} \mathrm{d}\hat{x}_n\, c(u_n, x_n, \hat{x}_n) \sum_{u_{\setminus\{j,n\}\cup\pi(j)}} \int_{\mathcal{X}_{\setminus\{j,n\}}} \mathrm{d}\hat{x}_{\setminus\{j,n\}}\, p(u_{\setminus j}, \hat{x}_{\setminus j}, x | u_j; \gamma_{\setminus j}^*) \\
&= \sum_{n\in an(j)\setminus\pi(j)} \sum_{u_n} \int_{\mathcal{X}_n} \mathrm{d}\hat{x}_n\, c(u_n, x_n, \hat{x}_n) p(x) p(u_n, \hat{x}_n | u_{\pi(j)}, u_j, x; \gamma_{\setminus j}^*) p(u_{\pi(j)} | u_j, x; \gamma_{\setminus j}^*) \\
&= \sum_{n\in an(j)\setminus\pi(j)} \sum_{u_n} \int_{\mathcal{X}_n} \mathrm{d}\hat{x}_n\, c(u_n, x_n, \hat{x}_n) p(x) p(u_n, \hat{x}_n | x; \gamma_{\setminus j}^*) p(u_{\pi(j)} | x; \gamma_{\setminus j}^*) \\
&= p(x) p(u_{\pi(j)} | x; \gamma_{\setminus j}^*) \sum_{n\in an(j)\setminus\pi(j)} \sum_{u_n} \int_{\mathcal{X}_n} \mathrm{d}\hat{x}_n\, c(u_n, x_n, \hat{x}_n) p(u_n, \hat{x}_n | x; \gamma_{\setminus j}^*)
\end{aligned}$$

revealing that it has no contribution on the optimization regarding the person-by-person optimal rule of node j either.

Therefore

$$\theta_j^*(u_j, \hat{x}_j, x, u_{\pi(j)}) \propto$$
$$p(x) p(u_{\pi(j)} | x; \gamma_{\setminus j}^*) \left[c(u_j, x_j, \hat{x}_j) + \sum_{m\in de(j)} \sum_{u_m} \int_{\mathcal{X}_m} \mathrm{d}\hat{x}_m\, c(u_m, x_m, \hat{x}_m) p(u_m, \hat{x}_m | x, u_j; \gamma_{\setminus j}^*) \right]$$

holds and under the measurement locality assumption, Eq.(18) easily yields

$$\phi_j^*(u_j, \hat{x}_j, x_j, u_{\pi(j)}) \propto$$
$$p(x_j) p(u_{\pi(j)} | x_j; \gamma_{\setminus j}^*) \left[c(u_j, x_j, \hat{x}_j) + \sum_{m\in de(j)} \sum_{u_m} \int_{\mathcal{X}_m} \mathrm{d}\hat{x}_m\, c(u_m, x_m, \hat{x}_m) p(u_m, \hat{x}_m | x_j, u_j; \gamma_{\setminus j}^*) \right] \quad (95)$$

after marginalization.

Now that we have obtained the form in Eq.(20) it remains to show that $p(u_{\pi(j)} | x_j; \gamma_{\setminus j}^*)$ is equal to $P_j^*(u_{\pi(j)} | x_j)$ given by the forward likelihood recursion Eq.s(21) and (22) together with that the summation over descendants is equal to $C_j^*(u_j, x_j)$ given by the induced cost recursion Eq.s(23) and (24).

We start with a general term $p(u_{\pi(j)} | x; \gamma)$ determined by the strategy γ and fist note that the directed acyclic nature together with the online processing in accordance with the forward ordering, $u_{\pi(j)}$ received from parents depend on $\gamma_{an(j)}$ and $x_{an(j)}$ yielding the equivalence $p(u_{\pi(j)} | x_{an(j)}; \gamma_{an(j)}^*) \equiv p(u_{\pi(j)} | x; \gamma^*)$ (Figure 8). In addition, starting with parentless nodes for which $p(u_{\pi(j)} | x; \gamma^*) = 1$ the following recursion holds where we denote by $u_{\pi^2(j)}$ the

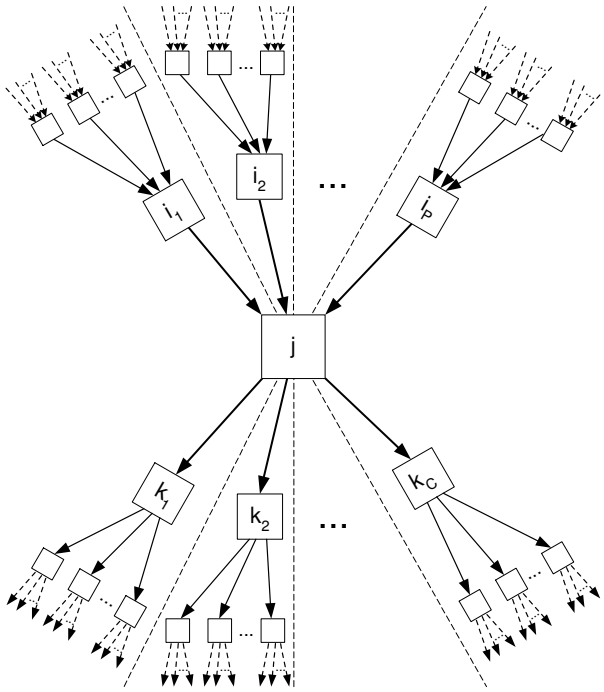


Fig. 8. A polytree from the viewpoint of node j: The parent nodes $i_1, i_2, ..., i_P$ do not have a common ancestor and the child nodes $k_1, k_2, ..., k_C$ do not have a common descendant.

set of incoming messages to parents of node j:

$$\begin{aligned} p(u_{\pi(j)}|x_{an(j)};\gamma^*_{an(j)}) &= \sum_{u_{\pi^2(j)}} \int_{\mathcal{X}_{\pi(j)}} \mathrm{d}\hat{x}_{\pi(j)}\, p(u_{\pi^2(j)}, u_{\pi(j)}, \hat{x}_{\pi(j)}|x_{an(j)};\gamma^*_{an(j)}) \\ &= \sum_{u_{\pi^2(j)}} \int_{\mathcal{X}_{\pi(j)}} \mathrm{d}\hat{x}_{\pi(j)} p(u_{\pi^2(j)}|x_{an(j)};\gamma^*_{an(j)}) p(u_{\pi(j)}, \hat{x}_{\pi(j)}|u_{\pi^2(j)}, x_{an(j)};\gamma^*_{an(j)}) \\ &= \sum_{u_{\pi^2(j)}} p(u_{\pi^2(j)}|x_{an(j)\setminus\pi(j)};\gamma^*_{an(j)\setminus\pi(j)}) \prod_{i\in\pi(j)} \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i \sum_{u_i\setminus u_{i\to j}} p(u_i, \hat{x}_i|x_i, u_{\pi(i)};\gamma^*_i) \end{aligned} \tag{96}$$

In addition, the polytree topology implies that no two parents of node j share a common ancestor and moreover the sets of ancestors of parents of node j are disjoint. Hence

$$p(u_{\pi^2(j)}|x_{an(j)\setminus\pi(j)};\gamma^*_{an(j)\setminus\pi(j)}) = \prod_{i'\in\pi(j)} p(u_{\pi(i')}|x_{an(i')};\gamma^*_{an(i')})$$

and after substituting in Eq.(96) we obtain

$$\begin{aligned} p(u_{\pi(j)}|x_{an(j)};\gamma^*_{an(j)}) &= \sum_{u_{\pi^2(j)}} \prod_{i'\in\pi(j)} p(u_{\pi(i')}|x_{an(i')};\gamma^*_{an(i')}) \prod_{i\in\pi(j)} \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i \sum_{u_i\setminus u_{i\to j}} p(u_i, \hat{x}_i|x_i, u_{\pi(i)};\gamma^*_i) \\ &= \sum_{u_{\pi^2(j)}} \prod_{i\in\pi(j)} p(u_{\pi(i)}|x_{an(i)};\gamma^*_{an(i)}) \sum_{u_i\setminus u_{i\to j}} \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i p(u_i, \hat{x}_i|x_i, u_{\pi(i)};\gamma^*_i) \end{aligned}$$

$$= \prod_{i\in\pi(j)} \sum_{u_{\pi(i)}} \sum_{u_i \setminus u_{i\to j}} p(u_{\pi(i)}|x_{an(i)};\gamma^*_{an(i)}) \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i p(u_i,\hat{x}_i|x_i,u_{\pi(i)};\gamma^*_i) \tag{97}$$

Finally, in order to obtain $p(u_{\pi(j)}|x_j;\gamma^*_{an(j)})$ we multiply both sides of the above equation with $p(x_{an(j)}|x_j)$ and marginalize $X_{an(j)}$, i.e.,

$$\begin{aligned}
&p(u_{\pi(j)}|x_j;\gamma^*_{an(j)}) = \\
&= \int_{\mathcal{X}_{an(j)}} \mathrm{d}x_{an(j)}\, p(x_{an(j)}|x_j) \prod_{i\in\pi(j)} \sum_{u_{\pi(i)}} \sum_{u_i\setminus u_{i\to j}} p(u_{\pi(i)}|x_{an(i)};\gamma^*_{an(i)}) \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i p(u_i,\hat{x}_i|x_i,u_{\pi(i)};\gamma^*_i) \\
&= \int_{\mathcal{X}_{\pi(j)}} \mathrm{d}x_{\pi(j)} \int_{\mathcal{X}_{an(j)\setminus\pi(j)}} \mathrm{d}x_{an(j)\setminus\pi(j)}\, p(x_{\pi(j)}|x_j) p(x_{an(j)\setminus\pi(j)}|x_{\pi(j)},x_j) \prod_{i\in\pi(j)} p(u_{i\to j}|x_i,x_{an(i)};\gamma^*_i,\gamma^*_{an(i)}) \\
&= \int_{\mathcal{X}_{\pi(j)}} \mathrm{d}x_{\pi(j)}\, p(x_{\pi(j)}|x_j) \int_{\mathcal{X}_{an(j)\setminus\pi(j)}} \mathrm{d}x_{an(i)\setminus\pi(j)} \prod_{i'\in\pi(j)} p(x_{an(i')}|x_{\pi(j)},x_j,...) \prod_{i\in\pi(j)} p(u_{i\to j},\hat{x}_i|x_i,x_{an(i)};\gamma^*_i,\gamma^*_{an(i)}) \\
&= \int_{\mathcal{X}_{\pi(j)}} \mathrm{d}x_{\pi(j)}\, p(x_{\pi(j)}|x_j) \prod_{i\in\pi(j)} \int_{\mathcal{X}_{an(i)}} \mathrm{d}x_{an(i)} p(x_{an(i)}|x_i,...) p(u_{i\to j}|x_i,x_{an(i)};\gamma^*_i,\gamma^*_{an(i)}) \\
&= \int_{\mathcal{X}_{\pi(j)}} \mathrm{d}x_{\pi(j)}\, p(x_{\pi(j)}|x_j) \prod_{i\in\pi(j)} p(u_{i\to j}|x_i;\gamma^*_i,\gamma^*_{an(i)}) \\
&= \int_{\mathcal{X}_{\pi(j)}} \mathrm{d}x_{\pi(j)}\, p(x_{\pi(j)}|x_j) \prod_{i\in\pi(j)} \sum_{u_{\pi(i)}} \sum_{u_i\setminus u_{i\to j}} p(u_{\pi(i)}|x_i;\gamma^*_{an(i)}) \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i p(u_i,\hat{x}_i|x_i,u_{\pi(i)};\gamma^*_i)
\end{aligned}$$

which is nothing but Eq.(22) substituted in Eq.s(21), where $P^*_j(u_{\pi(j)}|x_j)$ represents $p(u_{\pi(j)}|x_j;\gamma^*_{an(j)})$ and $P^*_{i\to j}(u_{j\to i}|x_i)$ is identified as $p(u_{i\to j}|x_i;\gamma^*_i,\gamma^*_{an(i)})$. In the first step above, we exploit the chain rule and in the next step, we substitute the disjointness of ancestors of parents of node j due to the polytree topology while factorizing $p(x_{an(j)\setminus\pi(j)}|x_{\pi(j)},x_j)$. To show that the factorization holds, let the parents of node j be $\pi(j) \triangleq \{i_1,...,i_P\}$. Then applying the chain rule consecutively we obtain

$$\begin{aligned}
&p(x_{an(j)\setminus\pi(j)}|x_{\pi(j)},x_j) \\
&\quad = p(x_{an(i_1)}|x_{\pi(j)},x_j) p(x_{an(j)\setminus\pi(j)\cup an(i_1)}|x_{\pi(j)},x_j,x_{an(i_1)}) \\
&\quad = p(x_{an(i_1)}|x_{\pi(j)},x_j) p(x_{an(j)\setminus\pi(j)\cup an(i_1)\cup an(i_2)}|x_{\pi(j)},x_j,x_{an(i_1)},x_{an(i_2)}) p(x_{an(i_2)}|x_{\pi(j)},x_j,x_{an(i_1)}) \\
&\quad ... \\
&\quad = p(x_{an(i_1)}|x_{\pi(j)},x_j) p(x_{an(i_2)}|x_{\pi(j)},x_j,x_{an(i_1)}) ... p(x_{an(i_P)}|x_{\pi(j)},x_j,x_{an(i_1)},...,x_{an(i_{P-1})})
\end{aligned}$$

Moreover $(u_i,\hat{x}_i)$ are independent from any fields of X given $(X_i,X_{an(i)})$ with γ^*_i and $\gamma^*_{an(i)}$ determined.

Similar steps show that the cost recursion given by Eq.s(23) and (24) hold, i.e., Eq.(24) substituted in Eq.s(23) is equal to summation over $m \in de(j)$ in Eq.(95). Consider

$$\begin{aligned}
&\sum_{m\in de(j)} \sum_{u_m} \int_{\mathcal{X}_m} \mathrm{d}\hat{x}_m\, c(u_m,x_m,\hat{x}_m) p(u_m,\hat{x}_m|x_j,u_j;\gamma^*_{\setminus j}) \\
&= \sum_{k\in\chi(j)} \left[\int_{\mathcal{X}_k} \sum_{u_k} c(u_k,x_k,\hat{x}_k) p(u_k,\hat{x}_k|x_j,u_j;\gamma^*_{\setminus j}) + \sum_{m\in de(k)} \mathrm{d}\hat{x}_m \sum_{u_m} c(u_m,x_m,\hat{x}_m) p(u_m,\hat{x}_m|x_j,u_j;\gamma^*_{\setminus j}) \right]
\end{aligned}$$

and let the summation over $m \in de(j)$ be denoted by $C_j^*(u_j, x_j)$. Then the expression above becomes

$$\begin{aligned} & C_j^*(u_j, x_j) \\ &= \sum_{k \in \chi(j)} \left[\int_{\mathcal{X}_k} \mathrm{d}x_k \sum_{u_k} c(u_k, x_k, \hat{x}_k) p(u_k, \hat{x}_k | x_j, u_j; \gamma_{\setminus j}^*) + \int_{\mathcal{X}_k} \mathrm{d}x_k \sum_{u_k} C_k^*(u_k, x_k) p(u_k, \hat{x}_k | x_j, u_j; \gamma_{\setminus j}^*) \right] \\ &= \sum_{k \in \chi(j)} \int_{\mathcal{X}_k} \mathrm{d}x_k \sum_{u_k} \left[c(u_k, x_k, \hat{x}_k) + C_k^*(u_k, x_k) \right] p(u_k, \hat{x}_k | x_j, u_j; \gamma_{\setminus j}^*) \end{aligned} \tag{98}$$

where it is possible to extend the distribution $p(u_k, \hat{x}_k | x_j, u_j; \gamma_{\setminus j}^*)$ such that it is expressed in terms of the contributions of the rule local to node k, i.e.,

$$\begin{aligned} p(u_k, \hat{x}_k | x_j, u_j; \gamma_{\setminus j}^*) = & \int_{\mathcal{X}_{\pi(k)\setminus j}} \mathrm{d}x_{\pi(k)\setminus j} \int_{\mathcal{X}_k} \mathrm{d}x_k \sum_{u_{\pi(k)\setminus j}} p(x_{\pi(k)\setminus j}, x_k | x_j) p(u_{\pi(k)\setminus j} | x_{\pi(k)\setminus j}; \gamma_{\setminus j}^*) \times \\ & p(u_k, \hat{x}_k | x_j, x_{\pi(k)\setminus j}, x_k, u_j, u_{\pi(k)\setminus j}; \gamma_{\setminus j}^*) \end{aligned}$$

$$= \int_{\mathcal{X}_{\pi(k)\setminus j}} \mathrm{d}x_{\pi(k)\setminus j} \int_{\mathcal{X}_k} \mathrm{d}x_k \sum_{u_{\pi(k)\setminus j}} p(x_{\pi(k)\setminus j}, x_k | x_j) \prod_{m \in \pi(k)\setminus j} p(u_{m\to k} | x_m; \gamma_m^*, \gamma_{an(m)}^*) p(u_k, \hat{x}_k | x_k, u_{\pi(k)}; \gamma_k^*) \tag{99}$$

where we identify $p(u_{m\to k} | x_m; \gamma_m^*, \gamma_{an(m)}^*)$ as $P_{m\to k}^*(u_{m\to k} | x_m)$ and substituted in Eq.(99) and Eq.(98) yields the cost recursion Eq.s(23) and (24). ■

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, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Proceedings of the 2001 IEEE International Conference on Acoustic, Speech and Signal Processing*, May 2001.

[2] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 40–50, March 2002.

[3] G. J. Pottie and W. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, May 2000.

[4] C.-Y. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, August 2003.

[5] M. Yarvis and W. Ye, *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, M. Ilyas and I. Mahgoub, Eds. CRC Press, 2005, chp. 13, Tiered Architectures in Sensor Networks.

[6] W. Su, E. Çayırcı, and Özgür B. Akan, *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, M. Ilyas and I. Mahgoub, Eds. CRC Press, 2005, chp. 16, Overview of Communication Protocols for Sensor Networks.

[7] B. Krishnamachari, *Networking Wireless Sensors*. Cambridge University Press, 2005.

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

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

[10] H. Wackernagel, *Multivariate Geostatistics*, 3rd ed. Springer, 2003.

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

[12] W.-M. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," *IEEE Transactions on communications*, vol. 41, no. 11, pp. 1602–1605, November 1993.

[13] J. A. Gubner, "Distributed estimation and quantization," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1456–1459, July 1993.

[14] P. Venkitasubramaniam, L. Tong, and A.Swami, "Quantization for maximin are in distributed estimation," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3596–3605, July 2007.

[15] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 2210–2219, June 2005.

[16] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part ii: Unknown probability density function," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2784 – 2796, July 2006.

[17] ——, "Bandwidth-constrained distributed estimation for wireless sensor networks-part i: Gaussian case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131 – 1143, March 2006.

[18] A. Ribeiro, I. D. S. ad Jin-Jun Xiao, G. B. Giannakis, and Z.-Q. Luo, *Wireless Sensor Networks: Signal Processing and Communication Perspectives*, A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, Eds. John Wiley &Sons, 2007.

[19] A. Dogandzic and K. Qiu, "Decentralized random-field estimation for sensor networks using quantized spatially correlated data and fusion center feedback," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 6069–6085, December 2008.

[20] Y. Wang, P. Ishwar, and V. Saligrama, "One-bit distributed sensing and coding for field estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4433–4445, September 2008.

[21] Y. Wang and P. Ishwar, "Distributed field estimation with randomly deployed, noisy, binary sensors," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1177–1189, March 2009.

[22] H. Zhang, J. Moura, and B. Krogh, "Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2383–2395, June 2009.

[23] J. Cortez, "Distributed kriged kalman filter for spatial estimation," *IEEE Transactions on Automatic Control*, vol. 54, no. 12, pp. 2816–2827, December 2009.

[24] M. Çetin, L. Chen, J. W. Fisher III, A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, "Distributed fusion in sensor networks: A graphical models perspective," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 42–55, July 2006.

[25] A. T. Ihler, J. W. Fisher III, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *Journal of Machine Learning Research*, vol. 6, pp. 905 – 936, May 2006.

[26] A. T. Ihler, "Inference in sensor networks: Graphical models and particle methods," Ph.D. dissertation, MIT, 2005.

[27] O. P. Kreidl, "Graphical models and message-passing algorithms for network-constrained decision problems," Ph.D. dissertation, MIT, 2008.

[28] O. P. Kreidl and A. S. Willsky, "An efficient message-passing algorithm for optimizing decentralized detection networks," in *Proceedings of the IEEE 45th Conference on Decision and Control*, December 2006, pp. 6776–6783.

[29] ——, "An efficient message-passing algorithm for optimizing decentralized detection networks," MIT Laboratory for Information and Decision Systems, Tech. Rep. LIDS-P-2726, 2007, iEEE Transactions on Automatic Control, To appear.

[30] ——, "Decentralized detection in undirected network topologies," in *Proceedings of the IEEE 14th Workshop on Statistical Signal Processing*. SSP '07, August 2007, p. 650=654.

[31] O. P. Kreidl and A. Willsky, "An efficient message-passing algorithm for optimizing decentralized detection networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 3, pp. 563–578, March 2010.

[32] M. Üney and M. Çetin, "An efficient monte carlo approach for optimizing communication constrained decentralized estimation networks," in *Proceedings of the* 17^{th} *Eurapean Signal Processing Conference*. EURASIP, EUSIPCO' 09, August 2009.

[33] ——, "An efficient monte carlo approach for optimizing decentralized estimation networks constrained by undirected topologies," in *Proceeedings of the* 15^{th} *IEEE Statistical Signal Processing Workshop*. IEEE, SSP' 09, August 2009.

[34] A. Pete, K. Pattipati, and D. L. Kleinmann, "Optimization of detection network structures with multiple event structures," *IEEE Transactions on Automatic Control*, vol. 39, no. 8, pp. 1702–1707, August 1994.

[35] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part i–fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, January 1997.

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

[37] Y.-C. Ho and K.-C. Chu, "Team decision theory and information structures in optimal control problems," *IEEE Transactions on Automatic Control*, vol. 29, no. 1, pp. 22 – 28, February 1972.

[38] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," MIT Laboratory for Information and Decision Systems, Tech. Rep. LIDS-P-2551, October 2002.

[39] E. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky, "Non-parametric belief propagation," in *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. CVPR' 03, June 2003.

[40] E. B. Sudderth, M. I. Mandel, W. T. Freeman, and A. S. Willsky, "Visual hand tracking using nonparametric belief propagation," in *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop*, vol. 12. CVPR' 04, July 2004.

[41] E. B. Sudderth, "Graphical models for visual object recognition and tracking," Ph.D. dissertation, MIT, 2006.

[42] N. Metropolis and S. Ulam, "The monte carlo method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, September 1949.

[43] N. Metropolis, A. W.Rosenbluth, M. N. Rosenbluth, A. H. Teller, and W. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087 – 1092, June 1953.

[44] J. Geweke, "Bayesian inference in econometric models using monte carlo integration," *Econometrica*, vol. 57, no. 6, pp. 1317–1339, November 1989.

[45] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. Springer, 2004.

[46] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, November 1984.

[47] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 140, pp. 398–409, June 1990.

[48] M. Abramowitz and I. A.Stegun, Eds., *Handbook of Mathematical Functions*. Dover Publications, Inc., 1965, page 376.

[49] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images," in *Advances in Neural and Information Processing Systems*. Conference on Neural Information Processing Systems (NIPS), Nov. 29-Dec.2 1999, pp. 855–861.

[50] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Applied and Computational Harmonic Analysis*, vol. 11, pp. 89–123, 2001.

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

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