Monte Carlo Optimization of Decentralized Estimation Networks Over Directed Acyclic Graphs Under Communication Constraints

Murat Üney, *Member, IEEE*, and Müjdat Çetin, *Member, IEEE*

***Abstract*—Motivated by the vision of sensor networks, we consider decentralized estimation networks over bandwidth-limited communication links, and are particularly interested in the tradeoff between the estimation accuracy and the cost of communications due to, e.g., energy consumption. We employ a class of in-network processing strategies that admits directed acyclic graph representations and yields a tractable Bayesian risk that comprises the cost of communications and estimation error penalty. This perspective captures a broad range of possibilities for processing under network constraints and enables a rigorous design problem in the form of constrained optimization. A similar scheme and the structures exhibited by the solutions have been previously studied 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 scheme involves 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 the in-network processing strategies and their optimization. The proposed Monte Carlo optimization procedure operates in a scalable and efficient fashion and, owing to the nonparametric 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*—Communication constrained inference, decentralized estimation, graphical models, in-network processing, message passing algorithms, Monte Carlo methods, random fields, wireless sensor networks.**

Nomenclature

$u_{\pi(j)}, u_j$	Incoming messages to node j from its parents $\pi(j)$ and outgoing messages from node j to its children $\chi(j)$.
$\mathcal{U}_{\pi(j)}, \mathcal{U}_j$	The sets of all possible incoming messages to node j and all possible outgoing messages from node j.
y_j	Observation of node j which is a random draw from the random variable Y_j.
$\gamma_j(y_j, u_{\pi(j)})$	Local rule of node j.
$\Gamma_j^{\mathcal{G}}$	Space of rules local to node j that are feasible in the network $\mathcal{G}$.
γ	In-network processing strategy as a concatenation of all local rules.
$\Gamma^{\mathcal{G}}$	Space of all feasible strategies over $\mathcal{G}$.
$c(u, x, \hat{x})$	Cost function penalizing the communication u and the pair $(x, \hat{x})$.
$J(\gamma)$	Bayesian risk of γ given $c(u, x, \hat{x})$.
γ^*	Person-by-person (pbp) optimal in-network processing strategy.
$c_j(u_j, x_j, \hat{x}_j)$	Cost function local to node j.
ϕ_j^*	Function characterizing the j^{th} pbp optimal local rule.
$P_j^*(u_{\pi(j)}, x_j)$	Incoming message likelihood of node j in the pbp optimal strategy.
$C_j^*(u_j, x_j)$	*Cost-to-go* function for node j in the pbp optimal strategy.
$P_{i\to j}^*(u_{i\to j}\|x_i)$	Forward likelihood message from node i to node j in the pbp optimal strategy.
$C_{k\to j}^*(u_{j\to k}, x_j)$	Backward cost message from node k to node j in the pbp optimal strategy.
$I_k^*(u_{\pi(k)}, x_k; \gamma_k^*)$	Total conditional cost local to node k in the pbp optimal strategy.
$c_j^d(x_j, \hat{x}_j)$	Estimation cost function local to node j.
$c_j^c(x_j, \hat{x}_j)$	Communication cost function of node j.
λ	Unit conversion constant; estimation penalty per unit communication cost.
$J_{d\|x_k, u_{\pi(k)}}^*$	Conditional estimation cost local to node k in the pbp optimal strategy.
$J_{c\|x_k, u_{\pi(k)}}^*$	Conditional cost due to node k's communication rule in the pbp optimal strategy.
$[.]^l$	In Algorithm 2 , the value taken by [.] in the l^{th} iteration.

Manuscript received November 29, 2010; revised April 14, 2011 and July 12, 2011; accepted July 15, 2011. Date of publication August 04, 2011; date of current version October 12, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiang-Gen Xia. 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.

The authors are with the Faculty of Engineering and Natural Sciences, Sabancı University, Orhanlı-Tuzla 34956 İstanbul, Turkey (e-mail: muratuney@sabanciuniv.edu; mcetin@sabanciuniv.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2163629

$x_j^{(m)}, y_j^{(p)}$	m^{th} element of the sample set S_{x_j} generated from $p(x_j)$ and p^{th} element of the sample set S_{y_j} generated from $p(y_j)$.
$\tilde{[.]}^{*^r}$	In Section IV, r-step approximation to $[.]^*$
$\tilde{[.]}^l$	In Algorithm 3, approximate evaluation of $[.]$ in the l^{th} iteration.

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 necessitates 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 [9, ch. 12] and geostatistics data [10] such as temperature measurements over a field [11, ch. 1].

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] and [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 fusion center. 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 a class 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 fusion center and the peripherals. However, the variable representing the decision on the partition selection does not provide conditional independence for the observations, and consequently exact fusion of the messages is not tractable and Monte Carlo approximations are employed. 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 which 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., [26, ch. 5]).

We consider a class of in-network processing strategies that is composed of local communication and computation rules and operates over a subset of all available links such that a directed acyclic topology 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 a 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 (DAG), a strategy is achieved by having all nodes produce outgoing messages to their children and an estimate of the random variables 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. Hence a rigorous problem setting for decentralized inference under communication constraints is obtained in the form of a constrained optimization problem in which the objective function is a Bayesian risk that penalizes both estimation errors and the transmissions, and the feasible set of strategies is constrained by the corresponding graph representation that captures the availability and the capacity of the links.

This class of decentralized strategies together with the structures exhibited by the solution have been recently studied in [27] (see also [28]) in the context of decentralized detection. 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 the particle representations and approximate computational schemes for the operators replace the original expressions. As a result, the iterative solution is 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 scheme we propose features scalability with the cardinality of the sample sets required and can produce results for any set of distributions provided that independent samples can be generated from the marginals.

In the next section, we define the problem in a constrained optimization setting, and then we present the Team Decision Theoretic investigation in Section III. In Section IV, we introduce our MC optimization framework for in-network processing strategies over DAGs, and in Section V, we demonstrate the aforementioned features through examples.[1] Finally, we provide some observations together with possible future directions, and conclude in Section VI.

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 we consider. This paradigm operates over DAGs for "network constrained *online* processing" of the set of collected observations and was previously studied in [28]. Subsequently, in Section II-B, 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.

We consider N sensor platforms dispersed over a region. 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 $u_{i\to j}$ without errors from the set of admissible symbols $\mathcal{U}_{i\to j}$ which is finite and the number of elements $|\mathcal{U}_{i\to j}|$ is in accordance with the link capacity capturing the bandwidth constraints.[2] A particular example of such a network can be seen in Fig. 3 given in Section V.

Each sensor platform is associated with a set of possibly multidimensional variables. For example, in a random field estimation scenario, these variables might be the temperature, humidity or the flow vector at some location, possibly at the position of the platform. We denote a concatenation of the variables associated with node j by X_j. Let us denote the denumerable set from which X_j takes values from by $\mathcal{X}_j$. The random variables to be estimated is the union of those associated with the platforms and can be represented with a concatenation $X = (X_1, X_2, \ldots, X_N)^T$, which takes values from $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_N$. Similarly, node j might be collecting a number of possibly multidimensional observations and a concatenation of these observations is denoted by Y_j. The set from which Y_j takes values from is denumerable and denoted by $\mathcal{Y}_j$. All observations collected by the network is given by the vector $Y = (Y_1, Y_2, \ldots, Y_N)^T$ which takes values from $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_N$.

The statistical model of the variables and the observations is given by the joint cumulative distribution function $P_{X,Y}(X, Y)$ with the density $p_{X,Y}(X, Y)$. Note that there are no restrictions on the dimensionality of the fields of X and Y, i.e., $\dim(\mathcal{X}_j), \dim(\mathcal{Y}_j) \geq 1$ for $j \in \mathcal{V}$. Therefore, X_j and Y_j can

[1] The preliminary results of the proposed scheme appear in [29].

[2] For example, it is possible to represent a link with capacity $\log_2 d_{ij}$ bits with $\mathcal{U}_{i\to j}$ such that $|\mathcal{U}_{i\to j}| = d_{ij} + 1$ where $0 \in \mathcal{U}_{i\to j}$ indicates no transmission and enables a message censoring or selective communication scheme. In [27] (and [28]), a channel model is accommodated to consider communication link errors. In addition, various transmission schemes such as "broadcast" and "peer-to-peer" are captured. Our setting falls into "peer-to-peer" type communication in this perspective. We assume that the links are error-free and do not employ a channel model.

accommodate multi-modal variables of multiple dimensions. Hence, this model enables a broad range of possibilities for decentralized inference.

A. In-Network Processing Paradigm Over DAGs

Consider a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Let $u_{\pi(j)}$ denote the incoming messages to node j from its parent 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 $\mathcal{U}_{\pi(j)} \triangleq \bigotimes_{i\in\pi(j)} \mathcal{U}_{i\to j}$, 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 $\mathcal{U}_j \triangleq \bigotimes_{k\in\chi(j)} \mathcal{U}_{j\to k}$. The cardinalities of $\mathcal{U}_{\pi(j)}$ and $\mathcal{U}_j$ can be found as $|\mathcal{U}_{\pi(j)}| = \prod_{i\in\pi(j)} |\mathcal{U}_{i\to j}|$ and $|\mathcal{U}_j| = \prod_{k\in\chi(j)} |\mathcal{U}_{j\to k}|$, respectively.

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-B. 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 $\mathcal{G}$. 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). The aggregation of local rules denoted by γ is called a strategy, i.e., $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_N)$, and takes values from the set of feasible strategies given by $\Gamma^{\mathcal{G}} = \bigotimes_{v\in\mathcal{V}} \Gamma_v^{\mathcal{G}}$. Considering the space of all possible estimators, i.e., $\Gamma \triangleq \{\gamma | \gamma : \mathcal{Y} \to \mathcal{X}\}$, it holds that $\Gamma^{\mathcal{G}} \subset \Gamma$. The set of all messages in the network is given by $u \triangleq \{u_{i\to j} | (i,j) \in \mathcal{E}\}$, and takes values from $\mathcal{U} \triangleq \bigotimes_{(i,j)\in\mathcal{E}} \mathcal{U}_{i\to j}$. The global view of this paradigm is illustrated in Fig. 1(b).

B. Design Problem in a Constrained Optimization Setting

It is possible to select a cost c for the network described by any graph $\mathcal{G}$ such that an estimation error penalty for the pair $(x, \hat{x})$ and a cost due to the corresponding set of messages in the network u are assigned, i.e., $c : \mathcal{U} \times \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. In addition, given $\gamma = (\gamma_1, \ldots, \gamma_N) \in \Gamma^{\mathcal{G}}$, the tuple $(U^T, \hat{X}^T)^T = \gamma(Y)$ is a random vector conditionally independent of X given Y, denoted by $(U^T, \hat{X}^T)^T \perp\!\!\!\perp X \,|\, Y$, and the density $p(u, \hat{x}|y)$ is

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

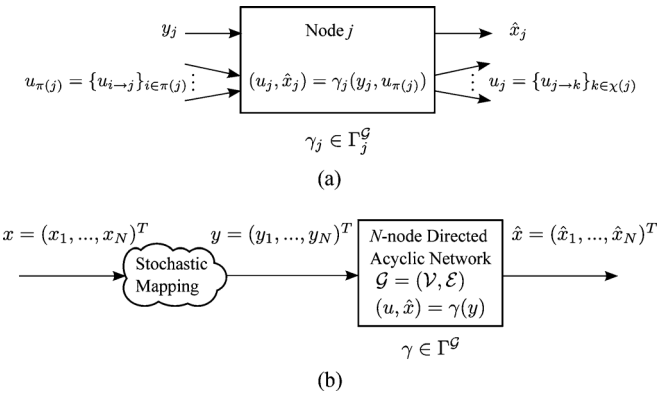

Fig. 1. Online processing scheme modeled with a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: (a) The viewpoint of node j in $\mathcal{G}$ which evaluates its local rule γ_j based on its measurement y_j as well as on the received messages $u_{\pi(j)}$ and produces an inference on the value of the random 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.

specified by γ and denoted by $p(u, \hat{x}|y; \gamma)$. Note that, by construction, considering the causal online processing in the DAG, it holds that

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

Consider a Bayesian risk $E\{c(u, x, \hat{x}); \gamma\}$. The distribution used in the expectation is specified by γ as well and the corresponding probability density function (pdf) is constructed using (1) as $p(u, \hat{x}, x; \gamma) = \int_{\mathcal{Y}} \mathrm{d}y p(u, \hat{x}|y; \gamma) p(y, x)$. Therefore, for any given strategy $\gamma \in \Gamma^{\mathcal{G}}$, there corresponds a Bayesian risk $J(\gamma) = E\{c(u, x, \hat{x}); \gamma\}$ 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

$$\begin{aligned} (\mathrm{P}): \quad & \min J(\gamma) \\ & \text{subject to } \gamma \in \Gamma^{\mathcal{G}}. \end{aligned} \tag{2}$$

It can be shown that if there exists an optimal strategy, then there exists an optimal deterministic strategy [30]. Therefore, it suffices to consider the deterministic local rule spaces for which case it is convenient to treat $p(u_j, \hat{x}_j | y_j, u_{\pi(j)}; \gamma_j)$ as a finite set of densities parameterized by u_j, i.e.,

$$p(u_j, \hat{x}_j | y_j, u_{\pi(j)}; \gamma_j) = p_{u_j}(\hat{x}_j | y_j, u_{\pi(j)}; \gamma_j) \tag{3}$$

$$p_{[\gamma_j(y_j, u_{\pi(j)})]_{\mathcal{U}_j}}(\hat{x}_j | y_j, u_{\pi(j)}; \gamma_j) = \delta\Big(\hat{x}_j - [\gamma_j(y_j, u_{\pi(j)})]_{\mathcal{X}_j}\Big) \tag{4}$$

where δ is the Dirac's delta distribution.[4] Hence, the local rule γ_j and the density family $p_{u_j}(\hat{x}_j | y_j, u_{\pi(j)}; \gamma_j)$ specify each other accordingly.

[4]We denote with $[.]_{\mathcal{S}}$ the element of its n-tuple argument that takes values from the set $\mathcal{S}$.

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 $\gamma_j^0 \in \Gamma_j^{\mathcal{G}}$ for $j = 1, 2, ..., N$; Choose $\varepsilon \in \mathbb{R}^+$; $l = 0$ ▷ Initialize

2: $l = l + 1$

3: For $j = N, N-1, \ldots, 1$ 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] and [13] as well as the bounded parameters estimation setting utilized in [14] and [17] through a noninformative prior within the framework above.

III. Team Decision Theoretic Investigation

Problem (P) in (2) is a typical team decision problem [31] and such problems are intractable in various settings, including conventional decentralized detection in which star-topologies are considered and $\mathcal{X}$ is finite [30]. 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 $\mathcal{V} \setminus \{j\}$.[5] Equivalently, a relaxation of (P) is to find a Nash equilibrium where no change in a single local rule yields a better objective value, i.e., one is interested in finding $\gamma^* = (\gamma_1^*, \ldots, \gamma_N^*)$ such that

$$\gamma_j^* = \arg\min_{\gamma_j \in \Gamma_j^{\mathcal{G}}} J(\gamma_j, \gamma_{\setminus j}^*) \tag{5}$$

for all $j \in \{1, 2, \ldots, N\}$. 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 pbp optimal strategies from the class of concern lie in a finitely parameterized subspace of $\Gamma^{\mathcal{G}}$ under certain conditions [27] and consequently a tractable iterative optimization algorithm is obtained. We adopt the elaborate investigation of Kreidl [28, ch. 3] for decentralized estimation under communication constraints and obtain a variational form for the pbp optimal local rules. These rules, unlike the pbp optimal local rules in the detection setting, are characterized through functions over denumerable domains and in general, do not yield any finite parameterization.

In principle, the Propositions regarding the pbp optimal estimation strategies given in this Section can be obtained from those in [27] by performing the marginalizations in the variables X_js and $\hat{X}_j$s through appropriate integrations under ideal channels and "peer-to-peer" transmission assumptions. In this respect, the proofs in this Section follow the same key steps with their detection counterparts. We also note that integrals over $\mathcal{X}_j$ or $\mathcal{Y}_j$ should be interpreted in accordance with the dimensionality of their domains.

The first condition that leads pbp optimal local rules to exhibit a useful structure is the conditional independence of the 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 [28] for Estimation): Consider (P) under Assumption 1. The j^{th} pbp optimal rule given by (5) reduces to

$$\gamma_j^*(y_j, u_{\pi(j)}) = \arg\min_{(u_j, \hat{x}_j) \in \mathcal{U}_j \times \mathcal{X}_j} \int_{\mathcal{X}} \mathrm{d}x\, p(y_j|x) \theta_j^*(u_j, \hat{x}_j, x; u_{\pi(j)}) \tag{6}$$

where

$$\theta_j^*(u_j, \hat{x}_j, x; u_{\pi(j)}) = p(x) \sum_{u_{\setminus j} \setminus u_{\pi(j)}} \int_{\mathcal{X}_{\setminus j}} \mathrm{d}\hat{x}_{\setminus j}\, c(u, \hat{x}, x) \prod_{i \neq j} \int_{\mathcal{Y}_i} \mathrm{d}y_i\, p(y_i|x) p(u_i, \hat{x}_i|y_i, u_{\pi(i)}; \gamma_i^*) \tag{7}$$

for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$ and $y_j \in \mathcal{Y}_j$ with nonzero 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 $\gamma_{\setminus j} = \gamma_{\setminus j}^*$, (1), (3), (4) and Assumption 1 together with the fact that if a pbp local rule exists, then a deterministic pbp local rule exists [30]. See [32] for a detailed proof. ■

Regarding Proposition 3.1 (and (6) 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.[6] 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 (6), 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 [27], is a function over a denumerable domain. In addition, it is useful to treat the right-hand side (RHS) of (7) 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 (6) which can also be treated as an operator given θ_j, i.e., $\gamma_j =$

[5]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 $\{x_1, x_2, \ldots, x_N\}$.

[6]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.

$\varsigma_j(\theta_j)$. 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

$$\theta_j^l = f_j(\theta_1^{l-1}, \ldots, \theta_{j-1}^{l-1}, \theta_{j+1}^l, \ldots, \theta_N^l) \tag{8}$$

for $j = 1, 2, \ldots, N$ where f_j denotes the composite operator (obtained after substituting $\varsigma_i(\theta_i)$ for all $i \in \backslash j$ in ψ_j). Note that, as a consequence of the fact that $\mathcal{X}$ is denumerable, the fixed point equations $\{\theta_j = f_j(\theta_{\backslash j})\}_{j\in\mathcal{V}}$ 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, although relatively straightforward numerical computations are involved when the joint density $p(x, y_1, \ldots, y_N)$ 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].

A. Pbp Optimal Strategies Over DAGs: Efficient Online Strategies

We continue with assumptions under which efficient online processing becomes possible [27]:

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

$$\gamma_j^*(Y_j, U_{\pi(j)}) = \underset{(u_j,\hat{x}_j)\in(\mathcal{U}_j\times\mathcal{X}_j)}{\arg\min} \int_{\mathcal{X}_j} \mathrm{d}x_j p(Y_j|x_j)\phi^*(u_j, \hat{x}_j, x_j; U_{\pi(j)}) \tag{9}$$

where

$$\phi_j^*(u_j, \hat{x}_j, x_j; u_{\pi(j)}) = \int_{x_{\backslash j}\in\mathcal{X}_{\backslash j}} \mathrm{d}x_{\backslash j}\theta_j^*(u_j, \hat{x}_j, x; u_{\pi(j)}). \tag{10}$$

Proof: Substitute $p(y_j|x) = p(y_j|x_j)$ in (6) 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.

B. Pbp Optimal Strategies Over DAGs: Efficient Offline Optimization

Corollary 3.2 provides an efficient online processing strategy. However, we do not have such efficiency for specifying the pbp optimal local rules since ϕ_j^* given by (10) 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.,

$$c(u, \hat{x}, x) = \sum_{j\in\mathcal{V}} c_j(u_j, \hat{x}_j, x_j). \tag{11}$$

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.

Proposition 3.3 (Proposition 3.2 in [28] for Estimation): Consider Problem (P) given in (2) such that X and $\hat{X}$ take values from a denumerable set $\mathcal{X}$. Under Assumptions 1–4, (9) applies with

$$\phi_j^*(u_j, \hat{x}_j, x_j; u_{\pi(j)}) \propto p(x_j)P_j^*(u_{\pi(j)}|x_j)\left[c_j(u_j, \hat{x}_j, x_j) + C_j^*(u_j, x_j)\right] \tag{12}$$

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

$$P_j^*(u_{\pi(j)}|x_j) = \begin{cases} 1, & \text{if } \pi(j) = \emptyset \\ \int\limits_{\mathcal{X}_{\pi(j)}} dx_{\pi(j)} p(x_{\pi(j)}|x_j) & \\ \quad \prod\limits_{i\in\pi(j)} P_{i\to j}^*(u_{i\to j}|x_i), & \text{otherwise} \end{cases} \tag{13}$$

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

$$P_{i\to j}^*(u_{i\to j}|x_i) = \sum_{u_{\chi(i)\backslash j}\in\mathcal{U}_{\chi(i)\backslash j}} \sum_{u_{\pi(i)}\in\mathcal{U}_{\pi(i)}} P_i^*(u_{\pi(i)}|x_i) \int\limits_{\mathcal{X}_i} \mathrm{d}\hat{x}_i \int\limits_{\mathcal{Y}_i} \mathrm{d}y_i\, p(u_i, \hat{x}_i|y_i, u_{\pi(i)}; \gamma_i^*)p(y_i|x_i) \tag{14}$$

and where $C_j^*(u_j, x_j)$ is the *cost-to-go* function which is added to the local cost and given by the backward recursion

$$C_j^*(u_j, x_j) = \begin{cases} 0, & \text{if } \chi(j) = \emptyset \\ \sum\limits_{k\in\chi(j)} C_{k\to j}^*(u_{j\to k}, x_j), & \text{otherwise} \end{cases} \tag{15}$$

with *backward* cost message regarding the influence of $k \in \chi(j)$ on j given by

$$C_{k\to j}^*(u_{j\to k}, x_j) = \int\limits_{\mathcal{X}_{\pi(k)\backslash j}} dx_{\pi(k)\backslash j} \int\limits_{\mathcal{X}_k} dx_k p(x_{\pi(k)\backslash j}, x_k|x_j) \sum_{u_{\pi(k)\backslash j}\in\mathcal{U}_{\pi(k)\backslash j}} \prod_{m\in\pi(k)\backslash j} P_{m\to k}^*(u_{m\to k}|x_m) I_k^*(u_{\pi(k)}, x_k; \gamma_k^*) \tag{16}$$

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

and where $I_k^*(u_{\pi(k)}, x_k; \gamma_k^*)$ is the *total conditional cost* of node k given by

$$I_k^*(u_{\pi(k)}, x_k; \gamma_k^*) = \int_{\mathcal{Y}_k} \mathrm{d}y_k \int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k \sum_{u_k\in\mathcal{U}_k} [c_k(u_k, \hat{x}_k, x_k) + C_k^*(u_k, x_k)] p(u_k, \hat{x}_k|y_k, u_{\pi(k)}; \gamma_k^*) p(y_k|x_k). \tag{17}$$

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 the local rules of the nodes that are ancestors of node 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.,

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

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 (12). 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 (13)–(14) and the additive structure of the expected costs in (15)–(17). A detailed proof is provided in [32, Appendix A]. ∎

Considering (13) and (14) 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 (15) and (16) 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 descendants of j for transmitting u_j, or cost-to-go function. This cost is added to the local cost $c_j(u_j, \hat{x}_j, x_j)$ in (12) which also penalizes the transmission cost. Also considering (9) and (10), 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 (8), it is possible to consider (13)–(17) as operators for any given (not necessarily 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

$$\phi_j = d_j(P_j, C_{\chi(j)\to j}) \tag{18}$$
$$P_j = f_j(P_{\pi(j)\to j}) \tag{19}$$
$$P_{j\to\chi(j)} = g_j(\phi_j, P_j) \tag{20}$$
$$C_{j\to\pi(j)} = h_j(\phi_j, P_{\pi(j)\to j}, C_{\chi(j)\to j}) \tag{21}$$

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 (12) and (15), (13), (14), and finally (16) and (17), 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-A, 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, in principle, through its message passing structure [27].

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 [28]. 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.,

$$c_j(u_j, x_j, \hat{x}_j) = c_j^d(x_j, \hat{x}_j) + \lambda c_j^c(u_j, x_j). \tag{22}$$

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 (9) is separated into two rules for estimation and communication as $\gamma_j^* = (\nu_j^*, \mu_j^*)$ given by

$$\hat{x}_j = \nu_j^*(y_j, u_{\pi(j)}) = \arg\min_{\hat{x}_j \in \mathcal{X}_j} \int_{x_j \in \mathcal{X}_j} dx_j p(x_j) p(y_j|x_j) \; P_j^*(u_{\pi(j)}|x_j) c_j^d(\hat{x}_j, x_j) \tag{23}$$

$$u_j = \mu_j^*(y_j, u_{\pi(j)}) = \arg\min_{u_j \in \mathcal{U}_j} \int_{x_j \in \mathcal{X}_j} dx_j p(x_j) p(y_j|x_j) \; P_j^*(u_{\pi(j)}|x_j) \left[\lambda c_j^c(x_j, u_j) + C_j^*(u_j, x_j)\right]. \tag{24}$$

Moreover, the corresponding density $p(u_j, \hat{x}_j|y_j, u_{\pi(j)}; \gamma_j^*)$ given by (3) takes the form

$$p(u_j, \hat{x}_j|y_j, u_{\pi(j)}; \gamma_j^*) = p(\hat{x}_j|y_j, u_{\pi(j)}; \nu_j^*) p(u_j|y_j, u_{\pi(j)}; \mu_j^*). \tag{25}$$

Proof: After substituting the separable local cost in (12) and (9), 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 independent denoted by $U_j \perp\!\!\!\perp \hat{X}_j \,|\, (Y_j, U_{\pi(j)})$ yielding (25) 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-square error (MSE) estimator. We obtain a closed form expression for the estimation rule regarding the variational form in (23) after differentiating with respect to $\hat{x}$ and equating the result to zero:

$$\hat{x}_j = \nu_j^*(Y_j, U_{\pi(j)}) = \frac{\int_{\mathcal{X}_j} \mathrm{d}x_j \, x_j p(x_j) p(Y_j|x_j) P_j^*(U_{\pi(j)}|x_j)}{\int_{\mathcal{X}_j} \mathrm{d}x_j \, p(x_j) p(Y_j|x_j) P_j^*(U_{\pi(j)}|x_j)}. \tag{26}$$

Note that, since $P_j^*(u_{\pi(j)}|x_j) = p(u_{\pi(j)}|x_j; \gamma_{an(j)}^*)$ and the conditional independence relation $U_{\pi(j)} \perp\!\!\!\perp Y_j \,|\, X_j$ holds yielding $p(x_j, y_j, u_{\pi(j)}) = p(x_j) p(y_j|x_j) p(u_{\pi(j)}|x_j)$, the denominator in (26) is nothing but $p(y_j, u_{\pi(j)}) = p(y_j, u_{\pi(j)}; \gamma_{an(j)}^*)$ and the estimator is given by

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

Hence, any selection of the local rules for ancestors affect the optimal estimation rule for node j through the likelihood $P_j^*(u_{\pi(j)}|x_j)$. Under this particular choice of the estimation cost, $u_{\pi(j)}$ is treated as another conditionally independent observation while utilizing the MSE estimator based on the posterior.

Corollary 3.6: Consider Proposition 3.3, if the local costs are separable, then $I_k^*(u_{\pi(k)}, x_k; \gamma_k^*)$ given by (17) takes the form

$$I_k^*(u_{\pi(k)}, x_k; \gamma_k^*) = J_{d|x_k, u_{\pi(k)}}^* + J_{c|x_k, u_{\pi(k)}}^* \tag{27}$$

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

$$J_{d|x_k, u_{\pi(k)}}^* = \int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k \, c_k^d(\hat{x}_k, x_k) p(\hat{x}_k|x_k, u_{\pi(k)}; \nu_k^*) \tag{28}$$

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 descendants given by the cost-to-go function

[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 $\lambda = \frac{(1-\alpha)}{\alpha}$ and $J(\gamma) \propto J'(\gamma)$ yielding a graceful degradation of the estimation performance with λ.

$C_k^*(u_k, x_k)$, as well as the transmission cost captured by $c_k^c(u_k, x_k)$, i.e.,

$$J^*_{c|x_k,u_{\pi(k)}} = \sum_{u_k \in \mathcal{U}_k} \left(\lambda c_k^c(u_k, x_k) + C_k^*(u_k, x_k)\right) p(u_k|x_k, u_{\pi(k)}; \mu_k^*). \tag{29}$$

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

$$p(\hat{x}_k|x_k, u_{\pi(k)}; \nu_k^*) = \int_{\mathcal{Y}_k} \mathrm{d}y_k \, p(\hat{x}_k|y_k, u_{\pi(k)}; \nu_k^*) p(y_k|x_k) \tag{30}$$

and the conditional probability mass function of the outgoing messages specified by μ_k^* is given by

$$p(u_k|x_k, u_{\pi(k)}; \mu_k^*) = \int_{\mathcal{Y}_k} \mathrm{d}y_k \, p(u_k|y_k, u_{\pi(k)}; \mu_k^*) p(y_k|x_k). \tag{31}$$

Proof: After substituting the separable local cost for node k given by (22) in (17) and rearranging terms

$$\begin{aligned} I_k^*(u_{\pi(k)}, x_k; \gamma_k^*) = & \int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k c_k^d(\hat{x}_k, x_k) \\ & \int_{\mathcal{Y}_k} \mathrm{d}y_k p(\hat{x}_k|y_k, u_{\pi(k)}; \nu_k^*) p(y_k|x_k) \\ & + \lambda \sum_{u_k \in \mathcal{U}_k} \left[\lambda c_k^c(u_k, x_k) + C_k^*(u_k, x_k)\right] \\ & \int_{\mathcal{Y}_k} \mathrm{d}y_k p(u_k|y_k, u_{\pi(k)}; \mu_k^*) p(y_k|x_k) \end{aligned} \tag{32}$$

is obtained. ∎

Therefore, under Assumptions 1–5, sufficient conditions of pbp optimality are provided by (12)–(16) together with (27)–(31) 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

$$J(\gamma^l) = \sum_{j \in \mathcal{V}} G_j(\gamma_j^l) \tag{33}$$

$$\begin{aligned} G_j(\gamma_j^l) = & \int_{\mathcal{X}_j} \mathrm{d}x_j p(x_j) \sum_{u_{\pi j} \in \mathcal{U}_{\pi j}} P_j^{l+1}(u_{\pi(j)}|x_j) \int_{\mathcal{Y}_j} \mathrm{d}y_j \int_{\mathcal{X}_j} \mathrm{d}\hat{x}_j \\ & \sum_{u_j \in \mathcal{U}_j} c_j(u_j, \hat{x}_j, x_j) p(u_j, \hat{x}_j|y_j, u_{\pi(j)}; \gamma_j^l) p(y_j|x_j). \end{aligned} \tag{34}$$

IV. MC Optimization Framework for In-Network Processing Strategies Over DAGs

In Sections III-A and III-B 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 (18)–(21) summarize the characterization of j^{th} pbp optimal rule given by (13)–(17), respectively, applied for local rules that are 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 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 continuous Markov random fields and has been the motivation for algorithms relying on particle representations together with approximate computational schemes, including nonparametric belief propagation [33], which has been successfully applied in a number of contexts including articulated visual object tracking [34].

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 [35], [36] and importance sampling [37], [38] such that independent samples generated from only the marginal distributions of X and Y are required, i.e.,

$$\begin{aligned} S_{x_j} \triangleq & \{x_j^{(1)}, x_j^{(2)}, \ldots, x_j^{(M_j)}\} \text{ such that} \\ & x_j^{(m)} \sim p(x_j) \text{ for } m = 1, 2, \ldots, M_j \end{aligned} \tag{35}$$

$$\begin{aligned} S_{y_j} \triangleq & \{y_j^{(1)}, y_j^{(2)}, \ldots, y_j^{(P_j)}\} \text{ such that} \\ & y_j^{(p)} \sim p(y_j) \text{ for } p = 1, 2, \ldots, P_j \end{aligned} \tag{36}$$

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.

We employ these sets in a way that the corresponding proposal densities (for importance sampling) are products of the corresponding marginals. This approach is advantageous for a number of reasons: First, we only need to know the marginal densities local to nodes, which can be computed using those densities that are already necessary in Algorithm 2, i.e., $p(x_{\pi(i)}, x_i)$ and $p(y_i|x_i)$ for all $i \in \mathcal{V}$. Therefore, sampling can be carried out locally. The second advantage is that, it is a relatively straightforward task to generate pseudorandom numbers from an arbitrary probability density function provided that the inverse of the corresponding cumulative distribution can be evaluated (see, e.g., [38, ch. 2]).

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 (9) and (12)). In Section IV-A we approximate the pbp optimal rule assuming that the factors in the RHS of (12) are known over their entire domain sets. In the second step we proceed with approximating to the incoming message likelihood (Section IV-B). In Section IV-C, the node-to-node terms, i.e., forward likelihood messages $P_{i\to j}^*$ from the parents $i \in \pi(j)$ and backward cost messages $C_{k\to j}^*$ from the children $k \in \chi(j)$, are approximated. 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.

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 (9) and (12) in the case that the remaining are fixed at the optimum $\gamma_{\setminus j} = \gamma^*_{\setminus j}$. After substituting (12) in (9) we obtain

$$\gamma_j^*(Y_j, U_{\pi(j)}) = \underset{(u_j,\hat{x}_j)\in(\mathcal{U}_j\times\mathcal{X}_j)}{\arg\min} R_j^*(u_j, \hat{x}_j; Y_j, U_{\pi(j)}) \tag{37}$$

$$R_j^*(u_j, \hat{x}_j; y_j, u_{\pi(j)}) = \int_{\mathcal{X}_j} \mathrm{d}x_j p(x_j)p(y_j|x_j)P_j^*(u_{\pi(j)}|x_j) \left[c_j(u_j, \hat{x}_j, x_j) + C_j^*(u_j, x_j)\right] \tag{38}$$

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 [27], $\mathcal{X}_j$ is a denumerable set and the RHS of (38) 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 cost-to-go function, i.e., $C_j^*(u_j, x_j)$, depend on the remaining local rules $\gamma^*_{\setminus j}$ (see Section III-B) and do not necessarily admit closed form expressions for arbitrary $\gamma_{\setminus j} \in \Gamma^{\mathcal{G}}_{\setminus j}$.

Suppose that it is possible to evaluate $P_j^*(u_{\pi(j)}|x_j)$ and $C_j^*(u_j, x_j)$ over their entire domains. The integral on the RHS of (38) still prevents R_j^* to be evaluated exactly, in general. Nevertheless, an approximation is possible through the classical Monte Carlo method given M independent samples (such as the set S_{x_j} given by (35)) generated from $p(x_j)$:

$$\tilde{R}_j^*(u_j, \hat{x}_j; y_j, y_{\pi(j)}) = \frac{1}{|S_{x_j}|} \sum_{x_j\in S_{x_j}} p(y_j|x_j)P_j^*(u_{\pi(j)}|x_j) \left[c_j(u_j, \hat{x}_j, x_j) + C_j^*(u_j, x_j)\right] \tag{39}$$

where tilde denotes that $\tilde{R}_j^*$ is 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. After we substitute $\tilde{R}_j^*$ in place of R_j^* in the variational form of γ_j^* given by (37), we obtain a local rule that 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 (37) 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 (39) splits trivially to two approximations, i.e., $\tilde{\nu}_j^{*\,1}$ and $\tilde{\mu}_j^{*\,1}$.

Example 4.1: Consider a separable cost as discussed in Corollary 3.4 with a quadratic estimation cost as in Example 3.5. Equation (39) substituted in (37) implies that the explicit solution for the quadratic estimation error given by (26) is approximated by

$$\hat{x}_j = \tilde{\nu}_j^{*\,1}(y_j, u_{\pi(j)}) = \frac{\left(\sum_{m=1}^{M} x_j^{(m)} p(y_j|x_j^{(m)}) P_j^*(u_{\pi(j)}|x_j^{(m)})\right)}{\left(\sum_{m=1}^{M} p(y_j|x_j^{(m)}) P_j^*(u_{\pi(j)}|x_j^{(m)})\right)}. \tag{40}$$

For the case, a similar approximation to the local communication rule given by (24) can be obtained as

$$u_j = \tilde{\mu}_j^{*\,1}(y_j, u_{\pi(j)}) = \underset{u_j\in\mathcal{U}_j}{\arg\min} \sum_{m=1}^{M} p\left(y_j|x_j^{(m)}\right) P_j^*\left(u_{\pi(j)}|x_j^{(m)}\right) \left[\lambda c_j^c\left(x_j^{(m)}, u_j\right) + C_j^*\left(u_j, x_j^{(m)}\right)\right]. \tag{41}$$

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 $m = 1, 2, \ldots, M$, 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 computation of the message likelihood function given by (13).

Let us consider (13) 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 one-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 $m = 1, 2, \ldots, M$, where $x_i^{(m)} \in S_{x_i}$ for $i \in \pi(j)$. These elements constitute a sample set $S_{\pi(j)} \triangleq \{x_{\pi(j)}^{(m)}\}$, 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., [38, ch. 3]) for P_j^* through the importance sampling density $\prod_{i\in\pi(j)} p(x_i)$. Then, the importance weights are given by

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

with the corresponding approximation

$$\tilde{P}_j^{*\,1}(u_{\pi(j)}|x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M} \omega_j^{(m)(m')}} \sum_{m'=1}^{M} \omega_j^{(m)(m')} \times \prod_{i\in\pi(j)} P^*_{i\to j}\left(u_{i\to j}|x_i^{(m')}\right) \tag{42}$$

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

Let us turn to the computation of the cost-to-go term $C_j^*(u_j, x_j)$ and consider (15) 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 $m = 1, 2, \ldots, M$ 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 $m = 1, 2, \ldots, M$ and for all $u_j \in \mathcal{U}_j$, can be computed exactly by simply substituting them in the RHS of (15).

As a result, given evaluations of the node-to-node terms $P_{i\to j}^*$ and $C_{k\to j}^*$ at sample points generated from the appropriate marginal distributions, node j can find the one-step approximate incoming message likelihood $\tilde{P}_j^{*\,1}$ given by (42) at all possible $(u_{\pi(j)}, x_j^{(m)})$ pairs. In addition, the cost-to-go function C_j^* given by (15) can easily be evaluated at all possible $(u_j, x_j^{(m)})$ pairs as mentioned before. Note that, these sets of possible pairs are exactly the sample points employed in the one-step approximate pbp optimal rule $\tilde{\gamma}_j^{*\,1}(y_j, u_{\pi(j)})$ defined in Section IV-A. Hence, a further approximation to the pbp optimal rule is obtained by substituting $\tilde{P}_j^{*\,1}$ in place of P_j^* in the RHS of the expression for $\tilde{R}_j^*$ (i.e., (39)) and then substituting the result in place of R_j^* in the variational form of the j^{th} pbp optimal rule given in (37). Let $\tilde{\gamma}_j^{*\,2}$ denote this approximate 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 $x_j^{(m)} \in S_{x_j}$. In this Section, we are concerned with approximating the evaluations of the forward likelihood message $P_{i\to j}^*$ given by (14) and the backward cost message $C_{k\to j}^*$ given by (16) at the sample sets.

First, we consider the parent nodes $i \in \pi(j)$ and consider evaluation of (14) 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 (3) and (4) for all $i \in \pi(i)$. The multiple integral term in (14), 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 $m = 1, 2, \ldots, M$, 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 density $p(y_i)$. Utilizing $y_i^{(p)} \in S_{y_i}$ and the importance weights given by

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

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

$$\tilde{p}(u_i|x_i^{(m)}, u_{\pi(i)}; \gamma_i^*) = \frac{1}{\sum_{p=1}^{P} \omega_i^{(m)(p)}} \sum_{p=1}^{P} \omega_i^{(m)(p)} \delta_{u_i, [\gamma_i^*(y_i^{(p)}, u_{\pi(i)})]_{\mathcal{U}_i}} \quad (43)$$

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 the forward likelihood message given by (14), 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 (43) we obtain

$$\tilde{P}_{i\to j}^*(u_{i\to j}|x_i^{(m)}) = \sum_{u_{\chi(i)\setminus j} \in U_{\chi(i)\setminus j}} \sum_{u_{\pi(i)} \in u_{\pi(i)}} P_i^*(u_{\pi(i)}|x_i^{(m)}) \times \tilde{p}(u_i|u_{\pi(i)}, x_i^{(m)}; \gamma_i^*) \quad (44)$$

for $m = 1, 2, \ldots, M$ 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 (42) with (44) and obtain a further step in the progressive approximations to γ_j^*.

The remaining term to consider is the cost-to-go function of j on the branch initiated with its child $k \in \chi(j)$, i.e., the backward cost message $C_{k\to j}^*(u_{j\to k}, x_j)$ given by (16). We proceed with approximating the evaluations of this function at all possible $(u_{j\to k}, x_j^{(m)})$ pairs such that $x_j^{(m)} \in S_{x_j}$ and $u_{j\to k} \in \mathcal{U}_{j\to k}$. With a similar reasoning, we utilize Monte Carlo methods on the RHS of the expression obtained by substituting the total conditional cost local to node k given by (17) in the backward cost message given by (16).

In the Appendix, we approximate the total conditional cost function evaluated at all possible $(u_{\pi(k)}, x_k^{(m)})$ for $u_{\pi(k)} \in \mathcal{U}_{\pi(k)}$ and $x_k^{(m)} \in S_{x_k}$ and obtain $\tilde{I}^*(u_{\pi(k)}, x_k^{(m)}; \gamma_k^*)$. Given this approximation, we consider the backward cost message given by (16) which further requires forward likelihoods 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 concatenation of elements from the usual sample sets local to $j' \in \pi(k)\setminus j$ is distributed according to the product of the corresponding marginals. In other words, let us define $x_{\pi(k)\setminus j}^{(m)} \triangleq (x_{j'}^{(m)})_{j'\in\pi(k)\setminus j}$. Then it holds that $x_{\pi(k)\setminus j}^{(m)} \sim \prod_{j'\in\pi(k)\setminus j} p(x_{j'})$ and an importance sampling approximation to (16) is possible through the importance density $p(x_k)\prod_{j'\in\pi(k)\setminus j} p(x_{j'})$. Having computed $\tilde{I}^*(u_{\pi(k)}, x_k^{(m)}; \gamma_k^*)$ and utilizing the usual sample sets local to nodes $j' \in \pi(k)\setminus j$ together with the importance sampling weights

$$\omega^{(m)(m')} = \frac{p\left(x_{\pi(k)\setminus j}^{(m')}, x_k^{(m')}|x_j^{(m)}\right)}{p\left(x_k^{(m')}\right) \prod_{j'\in\pi(k)\setminus j} p(x_{j'}^{(m')})}$$

we obtain

$$\tilde{C}^*_{k\to j}(u_{j\to k}, x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M} \omega^{(m)(m')}} \sum_{m'=1}^{M} \omega^{(m)(m')} \times \sum_{u_{\pi(k)\setminus j}} \prod_{j'\in\pi(k)\setminus j} P^*_{j'\to k}(u_{j'\to k}|x_{j'}^{(m')}) \tilde{I}^*(u_{\pi(k)}, x_k^{(m')}; \gamma_k^*) \tag{45}$$

for $m = 1, 2, \ldots, M$ and for all $u_{j\to k} \in \mathcal{U}_{j\to k}$ which, after substituting in place of $C^*_{k\to j}$ in the RHS of (15) for $\chi(j) \neq \emptyset$ yields $\tilde{C}^*_j$, i.e.,

$$\tilde{C}^*_j(u_j, x_j^{(m)}) = \sum_{k\in\chi(j)} \tilde{C}^*_{k\to j}(u_{j\to k}, x_j^{(m)}) \tag{46}$$

for $m = 1, 2, \ldots, M$ 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 (42), we obtain a further approximation to P^*_j given by

$$\tilde{P}_j^{*\,2}(u_{\pi(j)}|x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M} \omega_j^{(m)(m')}} \sum_{m'=1}^{M} \omega_j^{(m)(m')} \times \prod_{i\in\pi(j)} \tilde{P}^*_{i\to j}(u_{i\to j}|x_i^{(m')}) \tag{47}$$

for $m = 1, 2, \ldots, M$ and for all $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$. This approximation together with $\tilde{C}^*_{k\to j}$ given by (46) 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 Sections 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 optimal rules, i.e., $\gamma_{\setminus j} = \gamma^*_{\setminus j}$, we have constructed a local rule for node j, denoted by $\tilde{\gamma}_j^{*\,3}(y_j, u_{\pi(j)})$, such that it is an approximation to the optimal rule γ^*_j given by (37) and (38), following the progression

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

where S_{x_j} is given by (35),

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

where $\tilde{P}_j^{*\,1}$ is given by (42),

$$\tilde{\gamma}_j^{*\,3}(y_j, u_{\pi(j)}) = \operatorname*{arg\,min}_{(u_j,\hat{x}_j)\in(\mathcal{U}_j\times\mathcal{X}_j)} \sum_{x_j\in S_{x_j}} p(y_j|x_j) \tilde{P}_j^{*\,2}(u_{\pi(j)}|x_j) \times \left[c_j(u_j, \hat{x}_j, x_j) + \tilde{C}^*_j(u_j, x_j)\right] \tag{48}$$

where $\tilde{C}^*_j(u_j, x_j)$ and $\tilde{P}_j^{*\,2}$ are given by (46) and (47), 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 forward message likelihoods and the backward cost messages. Note that, we have not approximated $\gamma^*_{\setminus j}$ up to this point and assumed that it is exactly known.

On the other hand, given S_{x_j} and S_{y_j}, the approximation framework is valid for rules local to any node $j \in \mathcal{V}$: Owing to fusing the forward message likelihoods via importance sampling, approximations of the node-to-node terms given by (44) and (45) 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 RHSs of all expressions within the framework introduced in Sections IV-A–IV-C, as operators valid for any strategy $\gamma \in \Gamma^{\mathcal{G}}$ including those in the "approximating" form, e.g., $\{\tilde{\gamma}_j^{*\,3}\}_{j\in\mathcal{V}}$ where $\tilde{\gamma}_j^{*\,3}$ is given in (48). For the rest of this paper, an approximation to a function that appears in the local rules refers to its corresponding approximation in (48) 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

$$\begin{aligned}
\tilde{\phi}_j(S_{x_j}, \hat{x}_j) &= \tilde{d}_j(\tilde{P}_j(S_{x_j}), \tilde{C}_{\chi(j)\to j}) \\
\tilde{P}_j(S_{x_j}) &= \tilde{f}_j(\tilde{P}_{\pi(j)\to j}) \\
\tilde{P}_{j\to\chi(j)} &= \tilde{g}_j(\tilde{\phi}_j(S_{x_j}, \hat{x}_j), \tilde{P}_j(S_{x_j}))
\end{aligned}$$

where

$$\begin{aligned}
\tilde{C}_{j\to\pi(j)} &= \tilde{h}_j(\tilde{\phi}_j(S_{x_j}, \hat{x}_j), \tilde{P}_{\pi(j)\to j}, \tilde{C}_{\chi(j)\to j}) \\
\tilde{P}_j(S_{x_j}) &= \{(\tilde{P}_j(u_{\pi(j)}|x_j), u_{\pi(j)}, x_j) \\
&\qquad |u_{\pi(j)} \in \mathcal{U}_{\pi(j)} \wedge x_j \in S_{x_j}\} \\
\tilde{P}_{\pi(j)\to j} &= \{\tilde{P}_{i\to j}(S_{x_i})\}_{i\in\pi(j)} \\
\tilde{P}_{i\to j}(S_{x_i}) &= \{(\tilde{P}_{i\to j}(u_{i\to j}, x_i), u_{i\to j}, x_i) \\
&\qquad |u_{i\to j} \in \mathcal{U}_{i\to j} \wedge x_i \in S_{x_i}\} \\
\tilde{P}_{j\to\chi(j)} &= \{\tilde{P}_{j\to k}(S_{x_j})\}_{k\in\chi(j)} \\
\tilde{C}_{\chi(j)\to j} &= \{\tilde{C}_{k\to j}(S_{x_j})\}_{k\in\chi(j)}
\end{aligned}$$

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

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

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 since $\hat{x}_j$ is a free variable that can take values from $\mathcal{X}_j$.

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 3, 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

Algorithm 3 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

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

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

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

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

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

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

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 corresponding to an approximation to a person-by-person optimum.

The approximate computation of the expected cost required in the Check step of Algorithm 3, i.e., $\tilde{J}(\tilde{\gamma}^l)$, is performed through a Monte Carlo approximation $\tilde{G}_j(\tilde{\gamma}_j^l)$ to $G_j(\gamma_j^l)$ given by (34) using the usual sample sets, i.e., S_{x_j} and S_{y_j}, as

$$\tilde{G}_j(\tilde{\gamma}_j^l) = \frac{1}{M}\sum_{m=1}^{M}\sum_{u_{\pi(j)}\in\mathcal{U}_{\pi(j)}} \tilde{P}_j^{l+1}\left(u_{\pi(j)}|x_j^{(m)}\right)\frac{1}{\sum_{p=1}^{P}\omega_k^{(m)(p)}}\times$$

$$\sum_{p=1}^{P}\omega_k^{(m)(p)} c_j\left(\left[\tilde{\gamma}_j^l(y_j^{(p)}, u_{\pi(j)})\right]_{\mathcal{U}_j}, \left[\tilde{\gamma}_j^l(y_j^{(p)}, u_{\pi(j)})\right]_{\mathcal{X}_j}, x_j^{(m)}\right) \tag{49}$$

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

$$\tilde{G}_j\left(\tilde{\gamma}_j^l\right) = \frac{1}{M}\sum_{m=1}^{M}\sum_{u_{\pi(j)}\in\mathcal{U}_{\pi(j)}} \tilde{P}_j^{l+1}\left(u_{\pi(j)}|x_j^{(m)}\right)\times$$

$$\left[\tilde{J}^{l}_{d|x_j^{(m)}, u_{\pi(j)}} + \lambda\sum_{u_j\in\mathcal{U}_j} c_j^c\left(u_j, x_j^{(m)}\right)\tilde{p}\left(u_j|x_j^{(m)}, u_{\pi(j)}; \tilde{\mu}_j^l\right)\right] \tag{50}$$

and after distributing the multiplication in the RHS of the equation above and substituting in (33) in place of $G_j(\gamma_j^l)$, we obtain $\tilde{J}(\tilde{\gamma}^l) = \tilde{J}_d(\tilde{\gamma}^l) + \lambda\tilde{J}_c(\tilde{\gamma}^l)$.

Considering the separable cost function as discussed in Corollary 3.4 and a quadratic estimation cost as in Example 3.5, the approximate local estimation and communication rules at the jth node have complexities of $O(M)$ and $O\left(|\mathcal{U}_j|\,M\right)$ given an observation and incoming message pair $(y_j, u_{\pi(j)})$ where $|\mathcal{U}_j|$ is as defined in Section II-A. In the offline iterations, after receiving the forward likelihood messages as $|\mathcal{U}_{i\to j}| \times M$ arrays from the parents, the likelihood messages to the child nodes are computed in $O\left(M^2\,|\pi(j)|\,\left|\mathcal{U}_{\pi(j)}\right|\right)$ where $\left|\mathcal{U}_{\pi(j)}\right|$ is as defined in Section II-A as well (for a parentless node, this complexity is of $O\left(|\mathcal{U}_j|\,MP\right)$). After receiving the backward cost messages from the child nodes as $|\mathcal{U}_{j\to k}| \times M$ arrays, the cost messages to the parents are computed in $O\left(M\left|\mathcal{U}_{\pi(j)}\right|\left(M\,|\pi(j)| + P\,|\mathcal{U}_j|\right)\right)$ (for a childless node, this complexity is of $O\left(\left|\mathcal{U}_{\pi(j)}\right|MP\right)$). Note that the per node complexity is polynomial in the sizes of the sample sets. Owing to the message passing nature, a single forward–backward iteration of the optimization scales with the number of nodes in the sense that, given N nodes, if v is the node with highest computational demand bounded by $O(C(v))$, then a single forward–backward iteration is bounded by $O\left(NC(v)\right)$. Hence, the computational complexity increases only linearly with the number of nodes, making the algorithm scalable. We also note that $\left|\mathcal{U}_{\pi(j)}\right|$ and $|\mathcal{U}_j|$ grow combinatorially with the number of parents and number of children of node j, respectively. Therefore, nodes with the highest in-degree and/or out-degree bound the computational requirements of the iterations.

Note that $\{J(\gamma^l)|l = 0, 1, 2, \ldots\}$ obtained through Algorithm 2 is nonincreasing, whereas $\{\tilde{J}(\tilde{\gamma}^l)\}$ in Algorithm 3, being a MC 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 $\mathrm{err}[l]$ 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 nonincreasing with high probability can be obtained through an operator τ (Check step of Algorithm 3), investigation of which is beyond the scope of this work.

V. Examples

In this section, we demonstrate Algorithm 3 introduced in Section IV 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

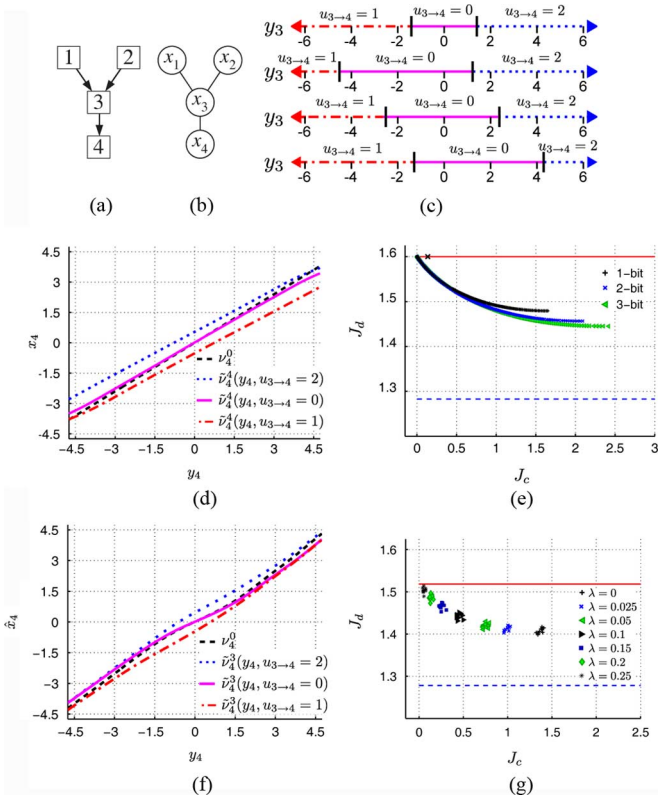

Fig. 2. (a) The DAG $\mathcal{G} = (\mathcal{V}, \mathcal{X})$, where $\mathcal{V} = \{1, 2, 3, 4\}$ and $\mathcal{E} = \{(1, 3), (2, 3), (3, 4)\}$. (b) The Markov random field (MRF) $\mathcal{G}_X$ subject to estimation by the decentralized estimation network in (a). (c) Communication rules for node 3: (top to bottom) the initial communication rule, i.e., $u_{3 \longrightarrow 4} = \mu_3^0(y_3, u_{\pi(3)})$ and illustrations of the converged communication rule for the Gaussian example for $\lambda = 0.1$ at the end of the 4^{th} step, specifically, $u_{3 \longrightarrow 4} = \mu_3^4(y_3, u_{\pi(3)})$ for $u_{\pi(3)} = \{2, 2\}, \{0, 0\}$ and $\{1, 1\}$, respectively. (d) 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. (e) The approximate performance points converged revealing the tradeoff together with the lower and upper bounds of the example Gaussian problem: Estimation performance measured in MSE for the optimum centralized and myopic rules correspond to the lower (blue dashed line) and the upper bound (red solid line), respectively. The estimation network in (a) is subject to optimization through Algorithm 3 starting with the initial rules given by (52) and (53) which achieve $(J_c(\gamma^0), J_d(\gamma^0))$ (black $\times$). 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. (f) For the Laplacian example; the converged estimation rule local to node 4 at the end of the 3^{rd} step, i.e., $\hat{x}_4 = \tilde{\nu}_4^3(y_4, u_{\pi(4)})$. (g) Approximate performance points achieved for the Laplacian case for ten sample sets and $\lambda = 0, 0.025, 0.05, \ldots, 0.25$.

$X = (X_1, X_2, X_3, X_4)^T$ is of concern and platform j is associated with X_j. In the first scenario, we consider a polytree online processing topology [Fig. 2(a)], a communication structure not covered by the star-topology paradigms (e.g., [13] and [19]), as well as stringent BW constraints such that $\mathcal{U}_{1 \longrightarrow 3} = \mathcal{U}_{2 \longrightarrow 3} = \mathcal{U}_{3 \longrightarrow 4} = \{0, 1, 2\}$. 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. 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 $(u_{1 \to 3}, \hat{x}_1) = \gamma_1(y_1)$ and $(u_{2 \to 3}, \hat{x}_2) = \gamma_2(y_2)$, respectively. Upon receiving these messages and measuring $y_3 \in \mathbb{R}$ induced by X_3, node 3 evaluates its local rule $(u_{3 \to 4}, \hat{x}_3) = \gamma_3(y_3, u_{1 \to 3}, u_{2 \to 3})$, and similarly node 4 evaluates $\hat{x}_4 = \gamma_4(y_4, u_{3 \to 4})$. The strategy $\gamma = (\gamma_1, \ldots, \gamma_4)$ is subject to design, which we perform through Algorithm 3.

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 $c_j(u_j, \hat{x}_j, x_j) = c_j^d(x_j, \hat{x}_j) + \lambda c_j^c(u_j, x_j)$ where the communication cost is additive as $c_j^c(u_j, x_j) = \sum_{k \in \chi(j)} c_{j \to k}^c(u_{j \to k}, x_j)$ and where $c_{j \to k}^c(u_{j \to k})$ is the cost of transmitting the symbol $u_{j \to k}$ on the link $(j, k) \in \mathcal{E}$ 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. 2(b). The covariance matrix is given by

$$\mathbf{C}_X = \begin{bmatrix} 2 & 1.125 & 1.5 & 1.125 \\ 1.125 & 2 & 1.5 & 1.125 \\ 1.5 & 1.5 & 2 & 1.5 \\ 1.125 & 1.125 & 1.5 & 2 \end{bmatrix} \tag{51}$$

which conforms with the Markov properties of $\mathcal{G}_X$. Although the communication structure of the decentralized estimation network is not related with the Markov random field representation of X and Algorithm 3 would produce results for any choice, for the sake of simplicity we selected the graph in Fig. 2(b) as the undirected counterpart of that in Fig. 2(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 6 dB.

Since separable local cost functions are utilized, the pbp optimal rules are also split into estimation and communications functions given by (23) and (24), respectively. Moreover, owing to the selection of c_j^d as the squared error estimation penalty, the local estimation rules take the form given in (26). 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

$$\nu_j^0(y_j, u_{\pi(j)}) = \int_{-\infty}^{\infty} \mathrm{d}x_j x_j p(x_j|y_j). \tag{52}$$

2) The initial communication rule of node j that is not childless is a quantization of the observation y_j, i.e.,

$$\mu_i^0(y_i, u_{\pi(i)}) = \begin{cases} 1, & y_i < -2\sigma_n \\ 0, & -2\sigma_n \leqslant y_i \leqslant 2\sigma_n \\ 2, & y_i > 2\sigma_n \end{cases} \tag{53}$$

regardless of the incoming messages.

Considering $J(\gamma) = J_d(\gamma) + \lambda J_c(\gamma)$ and pbp optimal strategies achieved by 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 30 000 samples for each S_{x_i} and S_{y_i} generated from $p(x_i)$ and $p(y_i)$, respectively, and use Algorithm 3 for varying λ from 0 with 0.001 steps. Example converged local communication and estimation rules are presented in Fig. 2(c) and (d) 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. 2(c) and the initial estimation rule illustrated by the black dashed line in Fig. 2(d) 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. 2(c)]. Now, let us turn to the estimation rule in Fig. 2(d). 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 behavior of $\tilde{v}_4^4$ for $u_{3\longrightarrow 4} = 2$ and $u_{3\longrightarrow 4} = 1$, respectively, in comparison with the initial rule as well as $\tilde{v}_4^4$ for $u_{3\longrightarrow 4} = 0$ in Fig. 2(d)], 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. 2(e) (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 (53) and the approximate performance points obtained are presented in Fig. 2(e) as well.[8] 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. Let us consider $(\tilde{J}_c, \tilde{J}_d)$ pairs for the 1-bit selective communication scheme, for $\lambda = 0$ (the transmission has no cost). The link use rate is approximately 1.65 bits, which is far less than the total capacity of 3 bits for the communication graph given in Fig. 2(a). This indicates that the information of receiving no messages is successfully maintained in this perspective. Moreover, the communication stops for $\lambda^* \approx 0.355$. Similarly, approximate points for 2-bit and 3-bit schemes [Fig. 2(e)] indicate that, if λ is small enough, we can achieve smaller MSE for the same total communication load as we increase the link capacities.

[8]For these experiments, we use the condition $\left|\tilde{J}(\tilde{\gamma}^{t-1}) - \tilde{J}(\tilde{\gamma}^{t})\right| < 1.0e-4$ in the `Check` step of Algorithm 3. The resulting average number of steps for convergence (within $\pm 3\sigma$) are 3.6 ± 1.5, 4.2 ± 2.0 and 4.1 ± 1.8 for $1, 2$ and 3-bit schemes, respectively. Please note that, for the graph given in Fig. 2(a), the complexity is determined by node 3 (Section IV-D). The time constant of forward messaging is fairly small compared to that for the backward messaging: In the 1-bit setting, node 3 computes the likelihood message to node 4 in 0.7713 sec., whereas the backward messages take approximately 200 sec. to be computed in a typical run using a 4-core PC with 8 GB of memory and nonoptimized MATLAB code. Under these conditions, one iteration is typically completed in approximately 272 sec.

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

$$p(x) = \frac{2}{(2\pi)^{\frac{d}{2}}|C_x|^{\frac{1}{2}}}\left(\frac{x^T C_x^{-1} x}{2}\right)^{\frac{1-d}{2}} K_{\frac{1-d}{2}}\left(\sqrt{2x^T C_x^{-1} x}\right) \tag{54}$$

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

Similar to that in the previous section, we assume the underlying communication structure described by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in Fig. 2(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 a MSL prior, even the centralized paradigm fails to provide a solution without employing numerical approximations.

For our case, we consider $X = (X_1, X_2, X_3, X_4)^T$ such that $p_X(x) = SL_4(\mathbf{C}_X)$ where $\mathbf{C}_X$ is given by (51) and we exploit the fact that the j^{th} marginal density of $SL_d(\mathbf{C}_X)$ is given by $SL_1([\mathbf{C}_X]_{j,j})$ and it is straightforward to generate samples from these marginals [40]. For the observations, although the marginal densities yield closed form expressions,[9] it is not easy to sample from this density since it does not yield a cumulative distribution function in closed form. Nevertheless, we can consider the mixture approximation $\sum_{x_j^{(m)} \in S_{x_j}} p\left(y_j | x_j^{(m)}\right) \approx$

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

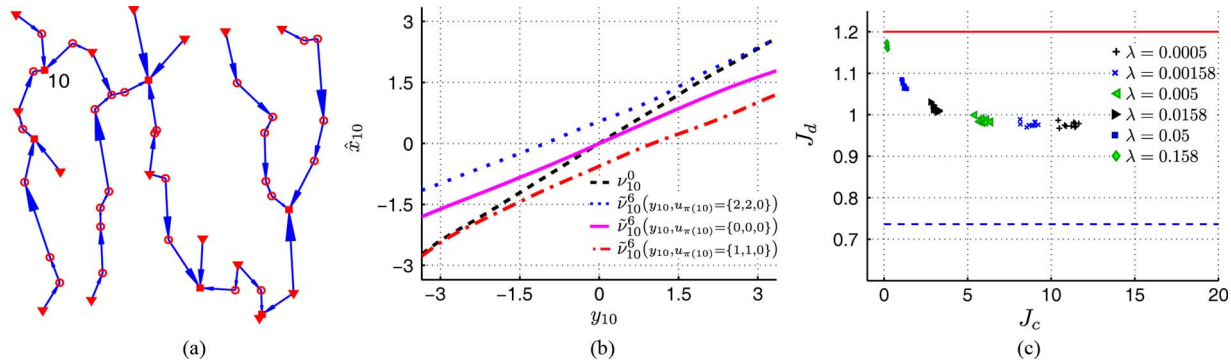

Fig. 3. (a) 50 randomly deployed nodes and the polytree generated from a spanning tree of the Gabriel graph of the deployment after randomly selecting six childless nodes; parentless and childless nodes are shown by red triangles and red squares (e.g., node 10), respectively. (b) Converged estimation rule local to (childless) node 10 for $\lambda = 0.005$ at the end of six iterations. (c) Approximate performance points of converged strategies for $\lambda = 0.0005, 0.00158, \ldots, 0.158$ and ten 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.

$p(y_j)$ where $S_{x_j} = \left\{x_j^{(1)}, x_j^{(2)}, \ldots, x_j^{(M)}\right\}$, and generate sufficient number of samples from $p\left(y_j|x_j^{(m)}\right)$ for each element of S_{x_j} so that the union of these samples are distributed by the mixture.

We generate $\left|S_{x_j}\right| = 3000$ samples from the prior marginals and $\left|S_{y_j}\right| = 45\,000$ samples from the aforementioned mixture densities by generating 15 samples from each component. We run Algorithm 3 for different choices of λ and for ten different sample sets. An example converged estimation rule is illustrated in Fig. 2(f), which is local to node 3 and obtained for $\lambda = 0.1$ after three 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. 2(f)] and is successfully represented within the MC framework. The asymptotic behaviors in the case that the measurement and the incoming message confirm each other are similar to that in the Gaussian example.

In Fig. 2(g), approximate performance points for the converged strategies based on the aforementioned ten sample sets are presented where the upper and lower bounds are the MSEs corresponding to the myopic and centralized rules, respectively. [10] In particular, for each value of λ, we generate results based on the ten sample sets. Collective results from such sample sets for a particular value of λ provide a sample-based approximation of the corresponding performance point on the Pareto-optimal curve. We observe that these sample-based results form clusters with reasonable variability, indicating their approximation quality. This level of variability can be expected since 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.

C. An Example With Large Graph

In this section, we consider a relatively large scale problem: 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 $X = (X_1, X_2, \ldots, X_{50})^T$ 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 Matérn covariance function which is commonly utilized in spatial data modeling [10] and given by

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

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 random deployment together with the polytree we generate by randomly selecting 6 childless nodes and employing Kruskal's algorithm on the Gabriel graph is given in Fig. 3(a).

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 merely provide information to their children based on the incoming messages from parents and the measurements they make. This is possible by simply selecting the estimation penalty as $c_j^d(x_j, \hat{x}_j) = (x_j - \hat{x}_j)^2$ if j is childless and zero otherwise. We consider a 1-bit selective communication scheme on each link and the communication cost considers the link use rate. Similar to the previous examples, the initial communication rules are quantization of the measurements and the childless nodes are initiated with the corresponding myopic estimation rules (for $\tau^2 + \sigma^2 = 1$, $\sigma_{n_j}^2 = 0.25\ \forall j \in \mathcal{V}$).

We employ Algorithm 3 for a geometrically increasing sequence $\lambda = 0.0005, 0.00158, \ldots, 0.158$ and for 10 different sample sets such that $\left|S_{x_j}\right| = 2000$ and $\left|S_{y_j}\right| = 30000$. An example converged estimation rule is illustrated in Fig. 3(b). We consider node 10 in Fig. 3(a); 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 behavior 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 midway 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. 3(c). Note that the myopic MSE for each platform is 0.2 yielding a total of 1.2,

[10] Another intricacy for this case is that the evaluation of the myopic and centralized strategies and the corresponding MSEs require numerical approximations for which we utilize MC methods as well.

whereas the centralized MSE (blue dashed line) is specified by the deployment (through $\mathbf{C}_x$). The MC framework successfully performs in large graph scenarios and similar to the example in Section V-B and the results given in Fig. 2(g), the performance points for different sample sets form clusters around the points from the Pareto-optimal curve they approximate and capture the tradeoff between estimation accuracy measured with MSE and the cost of communications in bits.

VI. 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 a class of *online* processing strategies over DAGs which is constituted of local rules operating in accordance with a (forward) message-passing structure on a DAG. This class provides a number of benefits compared with the conventional approaches in decentralized estimation including that it covers any association of the nodes with the variables that make up the global state and it is valid for any DAG presumably supported by the available set of links. Another important feature is that, under a Bayesian setting, it yields a rigorous design problem and a tractable *offline* strategy optimization procedure 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 tradeoff 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 over DAGs have been previously studied for decentralized detection [28] and hence our first contribution is the extension of these results for the estimation problem and a rephrasing of the offline optimization procedure which is composed of consecutive forward and backward message-passings.

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, i.e., the likelihoods of the incoming messages conditioned on the local rules of the ascendants and the expected cost induced on the descendants, as well as the pbp optimal local 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 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 V. 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.

One possible extension of this work, on which we have already obtained some preliminary results, is considering in-network strategies that are composed of two-stage local rules over undirected graphs (UGs). The family of strategies we considered also enable investigation of the two-stage strategies over UGs which render intertwined local star-graphs under certain assumptions and arguably better match some scenarios including the estimation of a random field. We have proposed a similar MC framework for this family yielding a similar optimization algorithm [41] which, together with the approximations presented in this work, can also potentially be applied for hybrid in-network processing strategies employing both families [42].

There are a number of issues left beyond the scope of this work. In contrast with the nonapproximated case, the iterative offline strategy optimization procedure does not yield a monotonically decreasing sequence of Bayesian risks. Investigation of a robust stopping condition remains as future work. Another possible extension is introduction of possible smoothing approaches through kernel methods. The IS estimate we use is known to be mildly biased [38, p. 95] and the analysis of the bias in our work remains an open issue. It might also be worthwhile to consider the problem of selecting the graph structure that yields the best pbp optimal strategy given an *a priori* distribution.

Appendix
Approximating the Total Conditional Cost

In this Appendix, we consider the total conditional cost local to node k, i.e., $I_k^*(u_{\pi(k)}, x_k; \gamma_k^*)$ given by (17) (in the context of Proposition 3.3), and approximate its evaluations at $x_k = x_k^{(m)}$ for all $x_k^{(m)} \in S_{x_k}$ and for all $u_{\pi(k)} \in \mathcal{U}_{\pi(k)}$. We assume that γ_k is fixed at the pbp optimal rule γ_k^*, and the density it specifies, i.e., $p(u_k, \hat{x}_k | y_k, u_{\pi(k)}; \gamma_k^*)$, is known. After substituting this density, which is given in (3) and (4) in Section II-B, into (17), we obtain

$$I_k^*(u_{\pi(k)}, x_k; \gamma_k^*) = \int_{\mathcal{Y}_k} \mathrm{d}y_k [c_k([\gamma_k^*(y_k, u_{\pi(k)})]_{\mathcal{U}_k}, [\gamma_k^*(y_k, u_{\pi(k)})]_{\mathcal{X}_k}, x_k) + C_k^*([\gamma_k^*(y_k, u_{\pi(k)})]_{\mathcal{U}_k}, x_k)] p(y_k | x_k), \tag{55}$$

the evaluation of which can be approximated at all $(u_{\pi(k)}, x_k^{(m)})$ pairs such that $x_k^{(m)} \in S_{x_k}$ and $u_{\pi(k)} \in \mathcal{U}_{\pi(k)}$ using the Importance Sampling method with the importance density $p(y_k)$. Assuming that the cost-to-go function local to node k, i.e., $C_k^*(u_k, x_k)$, is known for all possible $(u_k, x_k^{(m)})$ pairs and utilizing $y_k^{(p)} \in S_{y_k}$ together with the importance weights

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

we obtain the approximation given by

$$\tilde{I}_k^*\left(u_{\pi(k)}, x_k^{(m)}; \gamma_k^*\right) = \frac{1}{\sum_{p=1}^{P} \omega_k^{(m)(p)}} \sum_{p=1}^{P} \omega_k^{(m)(p)} \left[c_k\left(\left[\gamma_k^*(y_k^{(p)}, u_{\pi(k)})\right]_{\mathcal{U}_k}, \left[\gamma_k^*(y_k^{(p)}, u_{\pi(k)})\right]_{\mathcal{X}_k}, x_k^{(m)} \right) + C_k^*\left(\left[\gamma_k^*\left(y_k^{(p)}, u_{\pi(k)}\right)\right]_{\mathcal{U}_k}, x_k^{(m)} \right) \right] \quad (56)$$

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

In addition, if the separable cost assumption (Assumption 5 in Section III-B) holds, we consider Corollary 3.6 and find Importance Sampling approximations to the evaluations of the conditional estimation cost local to node k given by (28) and the conditional cost due to node k's communication rule given by (29) at $(u_{\pi(k)}, x_k^{(m)})$ with a similar reasoning we used above. These approximations to $J^*_{d|x_k^{(m)}, u_{\pi(k)}}$ and $J^*_{c|x_k^{(m)}, u_{\pi(k)}}$ are given by

$$\tilde{J}^*_{d|x_k^{(m)}, u_{\pi(k)}} = \frac{1}{\sum_{p=1}^{P} \omega_k^{(m)(p)}} \sum_{p=1}^{P} \omega_k^{(m)(p)} c_k^d(\nu_k^*(y_k^{(p)}, u_{\pi(k)}), x_k^{(m)})$$

$$\tilde{J}^*_{c|x_k^{(m)}, u_{\pi(k)}} = \sum_{u_k} \left(\lambda c_k^c(u_k, x_k^{(m)}) + C_k^*(u_k, x_k^{(m)}) \right) \tilde{p}(u_k|x_k^{(m)}, u_{\pi(k)}; \mu_k^*)$$

where $\tilde{p}(u_k|x_k^{(m)}, u_{\pi(k)}; \mu_k^*)$ is given by (43). Note that, $\tilde{J}^*_{d|x_k^{(m)}, u_{\pi(k)}} \approx J^*_{d|x_k^{(m)}, u_{\pi(k)}}$ and $\tilde{J}^*_{c|x_k^{(m)}, u_{\pi(k)}} \approx J^*_{c|x_k^{(m)}, u_{\pi(k)}}$ hold, and consequently, the approximation to the total conditional cost is obtained as $\tilde{I_k}^*(u_{\pi(k)}, x_k^{(m)}; \gamma_k^*) = \tilde{J}^*_{d|x_k^{(m)}, u_{\pi(k)}} + \tilde{J}^*_{c|x_k^{(m)}, u_{\pi(k)}}$.

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 *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (IEEE ICASSP)*, May 2001, vol. 4, pp. 2033–2036.

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

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

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

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

[6] W. Su, E. Çayirci, and Ö. B. Akan, *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, M. Ilyas and I. Mahgoub, Eds. Boca Raton, FL: CRC Press, 2005, ch. 16, Overview of Communication Protocols for Sensor Networks.

[7] B. Krishnamachari, *Networking Wireless Sensors*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[8] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Hawaii Int. Conf. Syst. Sci.*, Jan. 2000, DOI: 10.1109/HICSS.2000.926982 .

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

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

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

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

[13] J. A. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1456–1459, Jul. 1993.

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

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

[16] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks—Part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[17] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks—Part II: Unknown probability density function," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2784–2796, Jul. 2006.

[18] A. Ribeiro, I. D. Schizas , J.-J. 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. New York: Wiley, 2007, ch. 7, Distributed Estimation Under Bandwidth and Energy Constraints.

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

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

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

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

[23] J. Cortez, "Distributed Kriged Kalman filter for spatial estimation," *IEEE Trans. Autom. Control*, vol. 54, no. 12, pp. 2816–2827, Dec. 2009.

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

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

[26] A. T. Ihler, "Inference in sensor networks: Graphical models and particle methods," Ph.D. dissertation, Massachusetts Inst. of Technol., Cambridge, MA, 2005.
[27] O. P. Kreidl and A. Willsky, "An efficient message-passing algorithm for optimizing decentralized detection networks," *IEEE Trans. Autom. Control*, vol. 55, no. 3, pp. 563–578, Mar. 2010.
[28] O. P. Kreidl, "Graphical models and message-passing algorithms for network-constrained decision problems," Ph.D. dissertation, Massachusetts Inst. of Technol., Cambridge, MA, 2008.
[29] M. Üney and M. Çetin, "An efficient Monte Carlo approach for optimizing communication constrained decentralized estimation networks," in *Proc. 17th Eur. Signal Process. Conf. EURASIP (EUSIPCO)*, Glasgow, Scotland, U.K., Aug. 2009, pp. 1047–1050.
[30] J. N. Tsitsiklis, "Decentralized Detection," in *Advances in Statistical Signal Processing*. Greenwich, CT: JAI Press, 1993, pp. 297–344.
[31] Y.-C. Ho and K.-C. Chu, "Team decision theory and information structures in optimal control problems," *IEEE Trans. Autom. Control*, vol. 29, no. 1, pp. 22–28, Feb. 1972.
[32] M. Üney and M. Çetin, "Monte Carlo optimization approach for decentralized estimation networks under communication constraints," Sabanci Univ. ID:SU_FENS_2010/0007, 2010 [Online]. Available: http://research.sabanciuniv.edu/15985
[33] E. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," in *Proc. 2003 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. (CVPR)*, Madison, WI, Jun. 2003, pp. 605–612.
[34] E. B. Sudderth, "Graphical models for visual object recognition and tracking," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, 2006.
[35] N. Metropolis and S. Ulam, "The monte carlo method," *J. Amer. Stat. Assoc.*, vol. 44, no. 247, pp. 335–341, Sep. 1949.
[36] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and W. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, Jun. 1953.
[37] J. Geweke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 57, no. 6, pp. 1317–1339, Nov. 1989.
[38] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer, 2004.
[39] *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun, Eds. New York: Dover, 1965, p. 376.
[40] S. Kotz, T. J. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations*. Cambridge, MA: Birkhaüser, 2001.
[41] M. Üney and M. Çetin, "An efficient Monte Carlo approach for optimizing decentralized estimation networks constrained by undirected topologies," in *Proc. 15th IEEE Stat. Signal Process. Workshop (SSP)*, Aug. 2009, pp. 485–488.
[42] O. P. Kreidl and A. S. Willsky, "Decentralized detection with long-distance communication," in *Proc. Asilomar Conf. Signals, Syst., Comput. (CSSC)*, Oct. 2008, pp. 1352–1356.

Murat Üney (S'97–M'10) received the Ph.D. degree in electrical and electronics engineering from Middle East Technical University (ODTÜ), Ankara, Turkey, in 2009.

He was with the National Research Institute for Electronics and Cryptology (UEKAE), Ankara, between 2004 and 2007. Since 2007, he has been a member of the Signal Processing and Information Systems (SPIS) Laboratory within the Computer Vision and Pattern Analysis (VPA) Laboratory, Sabancı University, İstanbul. His research interests are in the broad scope of signal and information processing with emphases on statistical paradigms in distributed and multimodal problem settings and machine learning.

Müjdat Çetin (S'98–M'02) received the Ph.D. degree in electrical engineering from Boston University, Boston, MA, in 2001.

From 2001 to 2005, he was with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology (MIT), Cambridge, MA. Since September 2005, he has been a faculty member at Sabancı University, İstanbul, Turkey. His research interests include statistical signal and image processing, inverse problems, radar imaging, brain–computer interfaces, machine learning, computer vision, data fusion, wireless sensor networks, biomedical information processing, and sensor array signal processing.

Dr. Çetin has served as the Technical Program Co-Chair for the 2010 International Conference on Pattern Recognition and for the 2006 IEEE Turkish Conference on Signal Processing, Communications, and their Applications. He is currently an area editor for the *Journal of Advances in Information Fusion*, a Guest Editor for *Pattern Recognition Letters*, and a EURASIP Liaison Officer for Turkey. He has received several awards, including the 2010 IEEE Signal Processing Society Best Paper Award, the 2010 METU Parlar Foundation Research Incentive Award, the 2008 Turkish Academy of Sciences Distinguished Young Scientist Award, the 2007 Elsevier Signal Processing Journal Best Paper Award, and the 2006 TÜBİTAK Career Award.