STRUCTURED VARIATIONAL METHODS FOR DISTRIBUTED INFERENCE IN WIRELESS AD HOC AND SENSOR NETWORKS

Yanbing Zhang and *Huaiyu Dai*
Department of Electrical and Computer Engineering
NC State University
Raleigh, N.C. USA
{yzhang, Huaiyu_Dai}@ncsu.edu

Abstract –**In this paper, a variational message passing framework is proposed for Markov random fields, which is computationally more efficient and admits wider applicability compared to the belief propagation algorithm. Based on this framework, structured variational methods are explored to take advantage of both the simplicity of variational approximation (for inter-cluster processing) and the accuracy of exact inference (for intra-cluster processing). Its performance is elaborated on a Gaussian Markov random field, through both theoretical analysis and simulation results.**

Keywords: **Variational methods, distributed estimation, wireless networks**

I. INTRODUCTION

Distributed inference is the primary task of a wide range of wireless networking applications, like localization, tracking and time synchronization. In general, the objective of distributed inference is to compute marginal posterior probabilities through local interactions among nodes, given a set of observations and some underlying graph model. Message passing algorithms, like belief propagation (BP) [1], are attractive for distributed inference in wireless networks (WN) due to their energy efficiency, robustness and scalability [2]. However, BP and related algorithms are known not always to converge in general cyclic graphs, and computationally intractable when continuous variables are involved. Although sampling methods can be used for approximation, they are computationally intensive, stochastic in nature and difficult to analyze.

Variational methods [3], an alternative for approximate interference, involve the minimization of the Kullback–Leibler (KL) divergence between the target probability distribution and some "simpler" variational distribution. Being a deterministic approach, they are amenable to analysis, computationally efficient, and admit wide applications regarding the underlying models, whether acyclic or cyclic, discrete or continuous. Variational message passing (VMP) was studied in [5] for directed Bayesian networks. In this paper, we derive a variational message passing framework for Markov random fields (MRF), which arguably assumes certain advantages in modeling wireless networks. In particular, we formulate explicit message passing rules for distributions in the exponential family, which covers a large class of probabilistic models.

Distributed inference on a flat infrastructure enjoys the benefits of simplicity and uniformity in message formation, while higher accuracy and faster convergence can usually be achieved by organizing the whole network into a more structured one. However, structured variational approaches are mainly studied in the artificial intelligence area [7], and little consideration is given on their applications in real networks. In this work, we further investigate exploiting substructures of networks to improve variational methods in real systems. Thus the simplicity of variational methods and the high accuracy of exact inference algorithms can be exploited simultaneously. To our knowledge, this is the first work to explicitly apply the structured variational methods for distributed inference in wireless networks.

The rest part of this paper is organized as follows. Section II formulates the general variational method and derives its message passing form for MRF. Structured variational methods are proposed and analyzed in Section III. Section IV presents some supporting simulation results. Finally, concluding remarks are given in Section V.

II. VARIATIONAL MESSAGE PASSING IN MRF

A. Variational Method and Mean Field Approach

Assume $X=[X_1,...,X_N]^T$ is an N-dimension random vector and $y=[y_1,...,y_M]^T$ is the observation (M is not necessarily equal to N). The basic idea of variation methods is to approximate the posterior probability $P(X\mid y)$ with some variational distributions $Q(X)$ by minimizing the KL divergence between them:

$$KL(Q\,\|\,P)=\int_X Q(X)\log\frac{Q(X)}{P(X\mid y)}dX=-\mathbf{H}(Q)-\langle\log P(X\mid y)\rangle_{Q(X)}, \quad (1)$$

Where $\mathbf{H}(Q)$ is the entropy of $Q(X)$ and $\langle\ \rangle_{Q(X)}$ refers to expectation with respect to $Q(X)$. Clearly the best candidate is

$$Q^*(X)=\arg\max_{Q(X)}\left\{\langle\log P(X\mid y)\rangle_{Q(X)}+\mathbf{H}(Q)\right\}. \quad (2)$$

One simple and widely adopted form for variational distributions is to assume all variables are independent:

$$Q(X)=\prod_{i=1}^{N}Q_i(X_i), \quad (3)$$

which is also referred to as the mean field (MF) approach. This approach yields

$$\begin{aligned}&\langle\log P(X\mid y)\rangle_{Q(X)}+\mathbf{H}(Q)\\&=\int_X\prod_i Q_i(X_i)\log P(X\mid y)dX-\int_X\prod_i Q_i(X_i)\sum_i\log Q_i(X_i)dX \quad (4)\\&=\int_{X_k}Q_k(X_k)\langle\log P(X\mid y)\rangle_{\prod_{i\neq k}Q_i(X_i)}dX_k+\mathbf{H}(Q_k)+\sum_{i\neq k}\mathbf{H}(Q_i).\end{aligned}$$

It is easy to verify that maximizing the objective function in (4) with respect to the marginal $Q_k(X_k)$, $\forall k$, gives the standard Gibbs' distribution through the following set of fixed-point equations [3]

$$\log Q_k^*=\langle\log P(X\mid y)\rangle_{\prod_{i\neq k}Q_i^*(X_i)}+\text{constant terms},\ k=1,...,N. \quad (5)$$

B. Variational Message Passing in MRF

This work was supported in part by the US National Science Foundation under Grant CCF-0515164, CNS-0721815 and CCF-0830462.

Among well known graphical models, Markov random fields exhibit certain modeling convenience for wireless networks, as they can be conveniently mapped to real communication graphs [8]. A pairwise MRF[1] is an undirected graph (V,E) with maximum cliques of size 2, where each node $i \in V$ is associated with a random variable X_i. Define for each node a local potential function $\phi_i(X_i)$, and for each edge $(i,j) \in E$ a compatibility function $\psi_{ij}(X_i, X_j)$, the Hammersley-Clifford theorem [1] indicates that the posterior probability admits the following product form:

$$P(X \mid y) = \frac{1}{Z} \prod_{i \in V} \phi_i(X_i) \prod_{(i,j) \in E} \psi_{i,j}(X_i, X_j), \tag{6}$$

where Z is a normalization factor (called partition function in physics).

The exponential family includes a variety of commonly used distributions in practice, like Bernoulli, Poisson and Gaussian [4]. In the following, we assume that both the potential functions $\{\phi_i\}$ and the compatibility functions $\{\psi_{ij}\}$ take this form, given by

$$\phi_i(X_i, y) \propto \exp\left\{\boldsymbol{\theta}_i^T \boldsymbol{\eta}_i(X_i) + g_i(\boldsymbol{\theta}_i)\right\}, \tag{7}$$

$$\psi_{ij}(X_i, X_j) \propto \exp\left\{\boldsymbol{\theta}_{ij}^T \boldsymbol{\eta}_{ij}(X_i, X_j) + g_{ij}(\boldsymbol{\theta}_{ij})\right\}, \tag{8}$$

where $\boldsymbol{\theta}$'s and $\boldsymbol{\eta}$'s are usually referred to as the natural parameters and sufficient statistics, respectively.

To facilitate variational message passing, rearrange $\boldsymbol{\theta}_{ij}^T \boldsymbol{\eta}_{ij}(X_i, X_j)$ in terms of X_i:

$$\boldsymbol{\theta}_{ij}^T \boldsymbol{\eta}_{ij}(X_i, X_j) = (\boldsymbol{\theta}'_{ij})^T \boldsymbol{\eta}'_{ij}(X_i), \tag{9}$$

where $\boldsymbol{\theta}'_{ij}$ may be a function of X_j. Let $\tilde{\boldsymbol{\eta}}_i(X_i)$ be the *union* of sufficient statistics $\boldsymbol{\eta}_i(X_i)$ and $\boldsymbol{\eta}'_{ij}(X_i)$, then the corresponding terms in (7) and (8) can be rewritten as

$$\boldsymbol{\theta}_i^T \boldsymbol{\eta}_i(X_i) = \tilde{\boldsymbol{\theta}}_i^T \tilde{\boldsymbol{\eta}}_i(X_i), \tag{10}$$

$$\boldsymbol{\theta}_{ij}^T \boldsymbol{\eta}_{ij}(X_i, X_j) = (\boldsymbol{\theta}'_{ij})^T \boldsymbol{\eta}'_{ij}(X_i) = \tilde{\boldsymbol{\theta}}_{ij}^T \tilde{\boldsymbol{\eta}}_i(X_i). \tag{11}$$

We call $\{\tilde{\boldsymbol{\theta}}_i\}$ and $\{\tilde{\boldsymbol{\theta}}_{ij}\}$ the extended natural parameters and $\{\tilde{\boldsymbol{\eta}}_i(X_i)\}$ the extended sufficient statistics.

After some derivation (omitted in the interest of space), it can be shown that the optimal mean-field approximation Q_i^* is also a member of the exponential family with sufficient statistics $\tilde{\boldsymbol{\eta}}_i(X_i)$ and natural parameter vector

$$\boldsymbol{\theta}_i^* = \left\langle \tilde{\boldsymbol{\theta}}_i \right\rangle_{\prod_{k \neq i} Q_k^*(X_k)} + \sum_{j \in \Gamma_i} \left\langle \tilde{\boldsymbol{\theta}}_{ij} \right\rangle_{\prod_{k \neq i} Q_k^*(X_k)}, \tag{12}$$

where Γ_i stands for the neighboring node set of node i. Noting $\left\langle \tilde{\boldsymbol{\theta}}_i \right\rangle_{\prod_{k \neq i} Q_k(X_k)} = \tilde{\boldsymbol{\theta}}_i$ and $\left\langle \tilde{\boldsymbol{\theta}}_{ij} \right\rangle_{\prod_{k \neq i} Q_k(X_k)} = \left\langle \tilde{\boldsymbol{\theta}}_{ij} \right\rangle_{Q_j(X_j)}$, we can obtain the iterative variational message passing rules in MRF for the exponential family as:

- Messaging passing: $m_{j \to i}^{(n)} = \left\langle \tilde{\boldsymbol{\theta}}_{ij} \right\rangle_{Q_j^{(n-1)}(X_j)}$;
- Parameter updating: $\boldsymbol{\theta}_i^{(n)} = \tilde{\boldsymbol{\theta}}_i + \sum_{j \in \Gamma_i} m_{j \to i}^{(n)}$.

III. STRUCTURED VARIATIONAL METHODS FOR DISTRIBUTED INFERENCE IN WN

A. Structured Mean Field

Although attractive for its computational simplicity, the naive mean field approach may not yield sufficient accuracy or fast convergence due to the independence restriction on variational distributions. A natural idea for improvement is to integrate exact or more accurate probabilistic inference algorithms with the naive mean field method, if some probabilistically tractable substructures (clusters) can be identified. This approach is referred to as the structured mean field (SMF) method [6].

Given s clusters identified by some means, denoted as $C_1, \ldots, C_s$, the variational distribution Q for the SMF approach factors across the clusters as

$$Q = \prod_{i=1}^{s} Q_i(X_{C_i}). \tag{13}$$

The inter-cluster message updates resulting from the SMF approximation are analogous to the naive mean field method, which can be established through our previously proposed variational message passing framework. Intuitively we can interpret the SMF method as an MF approach over "mega variables" X_{C_i}. Within a cluster, any exact or more accurate inferencing algorithms can be applied, like the junction tree method or (loopy) belief propagation. As the cluster size increases, exact algorithms are performed on more nodes, which will result in a better approximation, while this also inevitably increases the computation burden. So it's important to choose appropriate cluster sizes to balance approximation accuracy and computation complexity.

SMF also requires some overhead for clustering, which can be done before the network setup and can be adjusted during network operation when necessary. We have designed a distributed clustering scheme suitable for wireless networks, explicitly considering nodes' energy and correlations, whose details are omitted here due to space limitations.

B. Distributed Inference in Structured Gaussian Markov Random Field

We elaborate the structured variational method in a Gaussian Markov random field, which is a widely adopted model for distributed inference in wireless networks.

Assume X is a Gaussian pairwise Markov random field and each node is only associated to a spatial component X_i of it. Each node makes a noisy linear observation

$$y_i = H_i x_i + \varepsilon_i,\ i = 1, \cdots, N, \tag{14}$$

where channel gain H_i is assumed known, and noise ε_i is Gaussian with zero mean and variance $R_i = \sigma^2$.

The prior marginal (assumed the same for all nodes) and joint correlation functions are formulated as[2]

$$p_i(x_i) = \mathcal{N}^{-1}(\mu_S / \sigma_S^2, 1/\sigma_S^2), \tag{15}$$

$$p_{ij}(x_i, x_j) = \mathcal{N}^{-1}(\Upsilon_{ij} \mu_S [1,1]^T, \Upsilon_{ij}),\ j \in \Gamma_i, \tag{16}$$

where

$$\Upsilon_{ij} = \frac{1}{\sigma_S^2 (1 - \rho_{ij}^2)} \begin{bmatrix} 1 & -\rho_{ij} \\ -\rho_{ij} & 1 \end{bmatrix}. \tag{17}$$

We adopt belief propagation for intra-cluster inference, for which clearly the messages and node beliefs are all Gaussian distributed.

[1] MRF with higher order cliques can always be converted into an equivalent pairwise MRF.

[2] Let $Z \sim \mathcal{N}(\mu, \Sigma)$ be a Gaussian random vector with mean μ and positive definite covariance Σ. One can define a new set of parameters (θ, Λ) by $\theta = \Sigma^{-1}\mu$, $\Lambda = \Sigma^{-1}$, and alternatively denote $Z \sim \mathcal{N}^{-1}(\theta, \Lambda)$.

Assuming the message $m_{ij}^{(n)}$ and the belief $b_i^{(n)}$ are parameterized by

$$m_{ij}^{(n)}(x) \sim \mathcal{N}^{-1}(\theta_{ij}^{(n)}, \Lambda_{ij}^{(n)}) \text{ and } b_i^{(n)}(x) \sim \mathcal{N}^{-1}(q_i^{(n)}, W_i^{(n)}),$$

we have obtained the updating rules as [8]

$$\begin{cases} \theta_{ij}^{(n)} = \dfrac{\rho_{ij}\left(\mu_i + \sum_{k\in\Gamma_i\setminus\{j\}} \theta_{ki}^{(n-1)} + \mu_S / \sigma_S^2(1+\rho_{ij})\right)}{1+\sigma_S^2(1-\rho_{ij}^2)\left(V_i + \sum_{k\in\Gamma_i\setminus\{j\}} \Lambda_{ki}^{(n-1)}\right)} + \dfrac{\mu_S}{\sigma_S^2(1+\rho_{ij})} \\ \Lambda_{ij}^{(n)} = \dfrac{V_i + \sum_{k\in\Gamma_i\setminus\{j\}} \Lambda_{ki}^{(n-1)} + 1/\sigma_S^2}{1+\sigma_S^2(1-\rho_{ij}^2)\left(V_i + \sum_{k\in\Gamma_i\setminus\{j\}} \Lambda_{ki}^{(n-1)}\right)}, \end{cases} \quad (18)$$

with

$$\begin{aligned} \mu_i &= H_i y_i / \sigma^2 + (1-|\Gamma_i|)\mu_S / \sigma_S^2 \\ V_i &= H_i^2 / \sigma^2 + (1-|\Gamma_i|) / \sigma_S^2, \end{aligned} \quad (19)$$

and

$$\begin{cases} q_i^{(n)} = \mu_i + \sum_{k\in\Gamma_i} \theta_{ki}^{(n)} \\ W_i^{(n)} = V_i + \sum_{k\in\Gamma_i} \Lambda_{ki}^{(n)}. \end{cases} \quad (20)$$

In the following, we discuss inter-cluster message updates. In a clustered graph, the posterior probability can be reformulated as

$$P(X \mid y) = \frac{1}{Z} \prod_{C_i \in C} \phi_{C_i}(X_{C_i}) \prod_{C_j \in \Gamma_{C_i}} \psi_{C_i, C_j}(X_{C_i}, X_{C_j}), \quad (21)$$

where $C = \{C_i\}$ denotes the entire cluster set, and Γ_{C_i} is the collection of Markov blanket clusters (MBC) of C_i. MBC are defined as the neighboring clusters that intersect with the Markov blanket (MB) of cluster C_i, $MB(C_i)$, which is the set of nodes outside of C_i but connected to some nodes in C_i. Fig. 1 gives a conceptual illustration for MB and MBC.

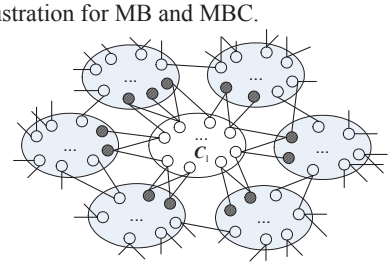

Fig. 1 Markov blanket (shaded nodes) and Markov blanket clusters (shaded clusters) for Cluster C_1. (reproduced Figure 1 in [7])

Regarding the clusters as "mega nodes" and following the line of [5], we can represent the extended natural parameters of the cluster potential and cluster compatibility functions as

$$\boldsymbol{\theta}_{C_i} = \{y_i / 2\sigma^2, i \in C_i\} \text{ and } \tilde{\boldsymbol{\theta}}_{C_i C_j} = \left\{-\rho_{ij} X_j / 2, i \in C_i, j \in \Gamma_{C_i}\right\}. \quad (22)$$

By substituting (22) into the variational message passing rules we derive in Section II.B, inter-cluster updating can be readily obtained for the "gateway" node i in the cluster C_i as

$$y_i^{(n)} = y_i^{(n-1)} - \sigma^2 \sum_{j \in MB(C_i)} \rho_{ij} \left(W_j^{(n-1)}\right)^{-1} q_j^{(n-1)}. \quad (23)$$

That is, gateway nodes use the estimates of their neighbors in the Markov blanket to "update" observations, and use these "new" observations for the next round intra-cluster inference.

C. Convergence Analysis

Combining inter- and intra-cluster updates, we can see the inverse variance updating in belief propagation is independent with the MF approach. Define function (c.f.(18))

$$F_{ij}(\Lambda) = \frac{V_i + \sum_{k\in\Gamma_i\setminus\{j\}} \Lambda_{ki} + 1/\sigma_S^2}{1+\sigma_S^2(1-\rho_{ij}^2)\left(V_i + \sum_{k\in\Gamma_i\setminus\{j\}} \Lambda_{ki}\right)}, \quad (24)$$

where $\Lambda = [\Lambda_{ij}] \in R^{2|E|\times 1}, (i,j) \in E$, then we have

Lemma 1: $F_{ij}(\Lambda)$ is continuous, monotonic and bounded for $\Lambda_{ij} > 0$, $\forall (i,j) \in E$.

Proof: Following the approach in [9], define function

$$f(x) = \frac{\alpha + x}{\beta + \gamma x} = \frac{1}{\gamma} + \left(\alpha - \frac{\beta}{\gamma}\right)\frac{1}{\beta + \gamma x}, \quad (25)$$

where $\alpha = V_i + 1/\sigma_S^2$, $\beta = 1 + \sigma_S^2(1-\rho_{ij}^2)V_i$, $\gamma = \sigma_S^2(1-\rho_{ij}^2)$. Since $1/\gamma > 0$ and $\alpha - \beta/\gamma = -\rho_{ij}^2 / [\sigma_S^2(1-\rho_{ij}^2)] < 0$, it's straightforward to verify that $f(x)$ is continuous, strictly increasing and bounded by $(0, 1/\gamma)$. This leads to the convergence of the inverse variance iteration. □

It is observed that the inverse variance iteration converges much faster than the message mean iteration. So we can allow the inverse variance iteration within every cluster to run first till the variance is sufficiently low. In the following, we analyze the convergence of the mean iteration under this assumption.

Without loss of generality, we assume $\mu_S = 0$, then the message mean iteration in Equation (18) can be rewritten with the conventional parameter pair (μ_{ij}, Σ_{ij}) as (c.f. footnote 2)

$$\mu_{ij}^{(n)} = G_{ij}(\mu^{(n-1)}) = \frac{\Sigma_{ij}^{(n)} \rho_{ij} \left(H_i y_i / \sigma^2 + \sum_{k\in N(i)\setminus\{j\}} \left(\Sigma_{ki}^{(n-1)}\right)^{-1} \mu_{ki}^{(n-1)}\right)}{1+\sigma_S^2(1-\rho_{ij}^2)\left(H_i^2 / \sigma^2 + \sum_{k\in N(i)\setminus\{j\}} \left(\Sigma_{ki}^{(n-1)}\right)^{-1}\right)}, \quad (26)$$

where $\mu^{(n-1)} = [\mu_{ij}^{(n-1)}] \in R^{2|E|\times 1}$, $\Sigma^{(n-1)} = [\Sigma_{ij}^{(n-1)}] \in R^{2|E|\times 1}$, $(i,j) \in E$.

Lemma 2: $G_{ij}(\mu)$ is a contraction mapping.

Proof:

Let $K_{ki\setminus\{j\}}^{(n-1)} = \dfrac{\sigma^2 \sigma_S^2 (1-\rho_{ij}^2)\left(\Sigma_{ki}^{(n-1)}\right)^{-1}}{\sigma^2 + \sigma_S^2 H_i^2 (1-\rho_{ij}^2)}$, $y_{ij}^{(n)} = \dfrac{\Sigma_{ij}^{(n)} \rho_{ij} H_i y_i}{\sigma^2 + \sigma_S^2 H_i^2 (1-\rho_{ij}^2)}$, and $\upsilon_{ij}^{(n)} = \dfrac{\Sigma_{ij}^{(n)} \rho_{ij}}{\sigma_S^2 (1-\rho_{ij}^2)}$, then Equation (26) can be reformulated to

$$\mu_{ij}^{(n)} = G_{ij}(\mu^{(n-1)}) = \frac{y_{ij}^{(n)} + \upsilon_{ij}^{(n)} \sum_{k\in N(i)\setminus j} K_{ki\setminus j}^{(n-1)} \mu_{ki}^{(n-1)}}{1 + \sum_{k\in N(i)\setminus j} K_{ki\setminus j}^{(n-1)}}. \quad (27)$$

Define $\gamma_{ij}^{(n-1)} = \dfrac{1}{1 + \sum_{k\in N(i)\setminus\{j\}} K_{ki\setminus\{j\}}^{(n-1)}}$, $D^{(n-1)} = \mathrm{diag}\left(\gamma^{(n-1)}\right) \in \mathrm{R}^{2|E|\times 2|E|}$, $V^{(n)} = \mathrm{diag}\left(\upsilon^{(n)}\right) \in \mathrm{R}^{2|E|\times 2|E|}$ and $y^{(n)} = [y_{ij}^{(n)}] \in \mathrm{R}^{2|E|\times 1}$. Further define a stochastic matrix $\hat{P}^{(n-1)} \in \mathrm{R}^{2|E|\times 2|E|}$ with

$$\hat{P}_{e,e'}^{(n-1)} = \begin{cases} \dfrac{K_{s(e')d(e')\setminus\{d(e)\}}^{(n-1)}}{\sum_{e':s(e')\in N(s(e))\setminus d(e)} K_{s(e')d(e')\setminus\{d(e)\}}^{(n-1)}}, & s(e) = d(e') \text{ but } s(e') \neq d(e) \\ 0, & \text{otherwise.} \end{cases} \quad (28)$$

The iteration (27) can be written in a vector-matrix form as

$$\mu^{(n)} = D^{(n-1)} y^{(n)} + V^{(n)} \left(I - D^{(n-1)}\right) \hat{P}^{(n-1)} \mu^{(n-1)}, \quad (29)$$

which leads to

$$\begin{aligned}\left\|G(\mu)-G(\mu')\right\|_\infty &= \left\|V^{(n)}\left(I-D^{(n-1)}\right)\hat{P}^{(n-1)}(\mu-\mu')\right\|_\infty \\ &\le \left\|V^{(n)}\left(I-D^{(n-1)}\right)(\mu-\mu')\right\|_\infty \\ &= \left\|V^{(n)}\left(I-D^{(n-1)}\right)\mathbf{1}\right\|_\infty \left\|(\mu-\mu')\right\|_\infty \\ &\le \eta\left\|(\mu-\mu')\right\|_\infty\end{aligned} \tag{30}$$

with $\eta<1$. The last inequality comes from the fact that for sufficiently large n, $\frac{\rho_{ij}}{\sigma_S^2(1-\rho_{ij}^2)}<\frac{1}{\Sigma_{ij}^{(n)}}$, i.e. $\upsilon_{ij}^{(n)}<1$. Thus, $G(\mu)$ is a maximum-norm contraction mapping and hence has a unique fixed point. This proves the convergence of the mean in Gaussian belief propagation. □

Now taking inter-cluster updating into account, the change in observations (23) will only reflect on $y^{(n)}$, which will be cancelled out in $G_{ij}(\mu)-G_{ij}(\mu')$. So Lemma 2 still can be applied. In conclusion, we have:

Theorem 1: In a Gaussian MRF, the structured variational method using Gaussian BP as the intra-cluster inference algorithm converges.

IV. SIMULATION RESULTS

We compare the performance of BP, MF and SMF on the same network. Note that BP and MF can be viewed as the two extremes for SMF, with one cluster, and clusters of size 1, respectively. We consider a Gaussian MRF estimation as discussed in Section III.B, and adopt a similar simulation setting as in [10], with 150 nodes being randomly distributed in a unit plane. The estimation mean square error (MSE) is used as the comparison metric, defined by

$$MSE=\left\|\hat{x}_m-\hat{x}\right\|_2/\left\|\hat{x}\right\|_2, \tag{31}$$

with $\hat{x}_m$ denoting the estimation vector at the mth iteration, and $\hat{x}$ being the exact estimation (the MMSE solution).

Fig. 2 compares the convergence rate of these three approaches, where the number of clusters is set to 65 and 15 respectively for SMF. The results verify that BP converges faster than the other two. But we also observe that even when the network is divided into very small clusters (65 clusters, corresponding to 2-3 nodes per cluster on average), SMF can still significantly speed up the convergence; in this case there is not much increase in computational complexity as compared to MF. A fairly large cluster size (15 clusters, corresponding to 10 nodes per cluster on average) achieves almost indistinguishable performance to that of BP.

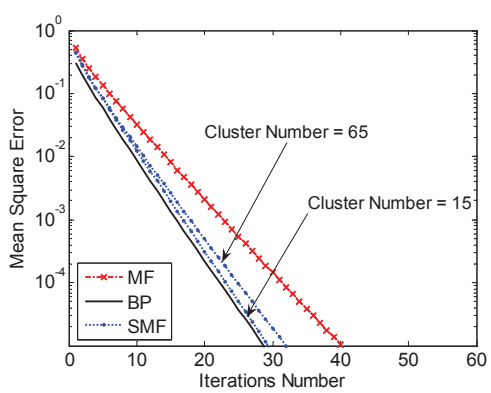

Fig. 2 Mean square error of estimation versus iteration number

Another important factor for wireless networks is the energy consumption. This comparison is illustrated in Fig. 3, where we assume for simplicity that the communication energy is proportional to the message complexity, and the computation energy is neglected. Note that compared to MF, BP exchanges more messages per round but converges faster. It is interesting to observe that for this simulation setting SMF consumes the least communication energy to obtain the same estimation accuracy, which indicates its potential superiority in practice.

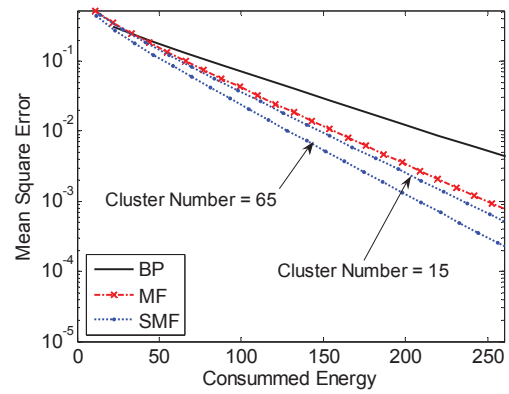

Fig. 3 Mean square error of estimation versus communication energy

Fig. 2 and Fig. 3 also indicate that an appropriate cluster size should be selected for SMF depending on applications, involving tradeoffs among estimation accuracy, convergence rate, computation complexity, and energy efficiency.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we develop a general variational message passing framework for distributed inference in Markov random fields. Structured variational methods are explored to achieve a nice tradeoff among various aspects of system performance and complexity. Our future work includes exploring its application in other scenarios (e.g. non-linear estimations) and more in-depth analysis.

REFERENCES

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[2] M. Cetin, L. Chen, J. Fisher, A. Ihler, R. Moses, M. J. Wainwright and A. S. Willsky. "Distributed fusion in sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, 2006.

[3] T. Jaakkola, "Tutorial on variational approximation methods," In *Advanced mean field methods: theory and practice*. MIT Press, 2000.

[4] M. J. Wainwright, and M. I. Jordan, "Graphical models, exponential families, and variational inference," UC Berkeley, Dept. of Statistics, Technical Report 649. September, 2003.

[5] J. Winn, "Variational message passing and its applications," Ph.D. Thesis, Department of Physics, University of Cambridge, 2003.

[6] L. Saul and M. Jordan, "Exploiting tractable substructures in intractable networks," In *Advances in Neural Information Processing Systems*, Vol. 8. The MIT Press, 1996.

[7] E.P. Xing, M.I. Jordan, and S. Russell, "A generalized mean field algorithm for variational inference in exponential families," In *Proceeding of 19th Annual Conference on Uncertainty in Artificial Intelligence* (UAI), 2003.

[8] H. Dai and Y. Zhang, "Consensus estimation via belief propagation," *2007 Conference on Information Sciences and Systems* (CISS), Baltimore, MD, March 2007.

[9] C. C. Moallemi and B. Van Roy, "Consensus Propagation," *IEEE Transactions on Information Theory*, Vol. 52, No. 11, pp. 4753-4766, 2006.

[10] V. Delouille, R. Neelamani, R. Baraniuk, "Robust Distributed Estimation using the Embedded Subgraphs Algorithm," *IEEE Trans. Signal Processing*, Vol. 54, No.8, 2006.