Proceedings of the 3rd IFAC Workshop on
Distributed Estimation and Control in Networked Systems
September 14-15, 2012. Santa Barbara, CA, USA

Solving the N-Consensus Problem: Combining Clustering and Synchronization

Stefan Gering, Volker Willert

Technische Universität Darmstadt, Institute of Automatic Control, Landgraf-Georg-Str. 4, 64283 Darmstadt, Germany, {sgering,vwillert}@rtr.tu-darmstadt.de

Abstract:
This article presents a state synchronization method within multi-agent systems upon multiple states. Based on their formation in state space, the agents decide on a clustering and synchronize their states within these clusters. The solution steps for this N-consensus problem, clustering and synchronization, may both be solved entirely in a decentral manner. This is achieved by means of a distributed Variational Bayes to describe the distribution of the agents' positions as a mixture of densities. The entire N-consensus problem is illustrated with graphical probabilistic models whose underlying potential is shown to be maximized when reaching the final N-consensus. An improvement of the overall convergence speed is achieved by a dynamical adaption of the distributed Variational Bayes, which leads to an intertwining of clustering and synchronization.

Keywords:
Autonomous mobile robots, consensus, distributed control, probabilistic models

1. INTRODUCTION

In mobile robotics, the synchronization of agent states within a multi-agent system (MAS) onto a single state is denoted as *consensus*. These states are often associated with positions in the spatial domain. To achieve this behavior within MAS by decentral computations, numerous distributed *consensus protocols* exist (e.g. (Saber and Murray, 2003), Moallemi and Van Roy (2006), Listmann et al. (2011)). These algorithms have in common that they solely aim to reach a synchronization upon a single state. Especially in MASs, which exhibit a large spatial distribution and which are supposed to solve tasks in smaller subgroups, it might be useful to synchronize the agents upon multiple different states depending on their group membership as well, which will be denoted as *N-consensus* in the sequel. Such an N-consensus is depicted in Fig. 1b for the MAS shown in Fig. 1a with its calculated group memberships. Again, the communication graph is shown by solid lines and the trajectories are depicted by dashed lines. The problem of synchronizing a MAS upon multiple states was already addressed, e.g. in Yu and Wang (2009), where a *multi-group consensus* was proposed. The drawback of the multi-group consensus protocol is that the number of groups as well as the agents' group memberships have to be known in advance. These parameters require global knowledge which contradicts the idea of a fully distributed protocol. Therefore, we present the N-consensus protocol that does neither require the number of groups nor the group memberships in advance. Instead, both will be the outcome of a decentral and distributed protocol. For this reason, the terminology *N-consensus* is used, emphazising that the number of groups is unknown a priori.

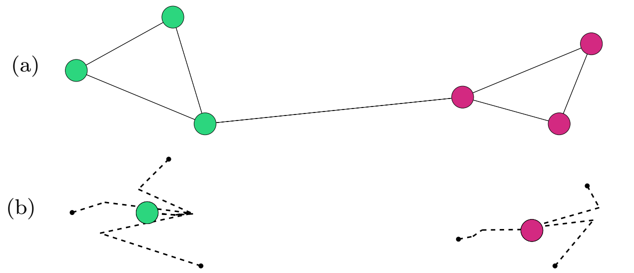

Fig. 1. (a) MAS with final cluster memberships shown by different colors and (b) N-consensus

Apparently, the N-consensus problem consists of a distributed clustering as well as a distributed synchronization depending on the clustering. Besides consensus protocols, there already exist methods for distributed clustering. The *affinity propagation* algorithm presented in Frey and Dueck (2007), for example, makes use of a message passing procedure. A distributed *K-means* algorithm discussed in Forero et al. (2008) as well as a distributed *Expectation Maximization (EM)* introduced in Gu (2008) may serve as other examples. The latter two approaches can be used in MAS with arbitrary connected communication graphs. Nevertheless, they suffer from the drawback that the number of clusters K has to be known in advance, which may only be determined when the MAS formation is known. This requirement of global knowledge appears to be inconvenient for a completely decentralized implementation.

In Safarinejadian et al. (2010), a distributed Variational Bayes is introduced, which on the one hand does not suffer from this drawback, because K is automatically determined during the iterative solution. On the other hand, this algorithm is limited to the special case of sensor networks with a ring-shaped topology. Because of

978-3-902823-22-9 /12/$20.00 © 2012 IFAC

10.3182/20120914-2-US-4030.00006

this, an application to MAS with arbitrary connected communication graphs is not possible.

In the following, a solution for the distributed calculation of the Variational Bayes in MAS with arbitrary connected communication topology is presented. In contrast to Safarinejadian et al. (2010), this allows an estimation of the distribution of the agents' positions, which forms the basis of an N-consensus within MAS – a problem which to the best of our knowledge has not been addressed before.

The paper is organized as follows: In Sec. 2, the N-consensus problem is formally stated. Sec. 3.1 then reviews the concept of the Variational Bayes for mixtures of Gaussians as described in Bishop (2006), which is then extended in Sec. 3.2 for fully distributed implementations. A further adaption of this concept in combination with an appropriate determination of control signals leads to the N-consensus protocol in Sec. 4. Then, Sec. 5 presents a way of graphically representing probabilistic dependencies occuring in the N-consensus problem and Sec. 6 illustrates the application of the framework for an example.

2. THE N-CONSENSUS PROBLEM

We consider MAS with N agents, whose communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is represented by a finite non-empty node set $\mathcal{V}(\mathcal{G})$ and an edge set $\mathcal{E}(\mathcal{G}) \subseteq \mathcal{V} \times \mathcal{V}$. The nodes $v_n \in \mathcal{V}(\mathcal{G})$ describe the agents and the undirected edges $e_n \in \mathcal{E}(\mathcal{G})$ represent communication links between them. The set of *neighbors* $\mathcal{N}(n)$ of node v_n contains all nodes v_m for which $(v_n, v_m) \in \mathcal{E}(\mathcal{G})$ holds. It is assumed that $\mathcal{G}$ is always connected. For simplicity, it is also assumed that the agents exhibit integrator dynamics

$$\dot{\boldsymbol{x}}_n = \boldsymbol{u}_n \tag{1}$$

with $\boldsymbol{x}_n(0) = \boldsymbol{x}_{n,0} \in \mathbb{R}^D$, where $\boldsymbol{x}_n$ describes the position of the n-th agent and $\boldsymbol{u}_n$ describes its actuator signal. Following Saber and Murray (2003), the choice of

$$\boldsymbol{u}_n(t) = \sum_{v_m \in \mathcal{N}(n)} \big(\boldsymbol{x}_m(t) - \boldsymbol{x}_n(t)\big) \tag{2}$$

leads to the convergence of the agent states towards the mean $\boldsymbol{x}_{\text{syn}} = \frac{1}{N} \cdot \sum_{n=1}^{N} \boldsymbol{x}_{n,0}$ of their initial positions and therefore onto a single synchronous state. Because an N-consensus shall be reached, the consensus protocol in (2) may not be applied directly. Instead, the agents first have to estimate their spatial distribution and come to a state synchronization upon their cluster centers $\boldsymbol{\mu}_k$ by means of an actuator signal

$$\boldsymbol{u}_n(t) = \left(\sum\nolimits_{k=1}^{K} z_{nk} \cdot \boldsymbol{\mu}_k\right) - \boldsymbol{x}_n(t) \tag{3}$$

or reference input

$$\boldsymbol{w}_n(t) = \boldsymbol{u}_n(t) + \boldsymbol{x}_n(t) = \sum\nolimits_{k=1}^{K} z_{nk} \cdot \boldsymbol{\mu}_k, \tag{4}$$

depending on the cluster membership $z_{nk} \in \{0, 1\}$.

3. DISTRIBUTED PROBABILISTIC CLUSTERING

3.1 Variational Bayes

In order to achieve a clustering among the agents, their positions $\boldsymbol{x}_n$ are interpreted as independently identically distributed stochastic variables and it is assumed that positions of the same cluster result from a multivariate normal distribution $\mathcal{N}\left(\boldsymbol{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)$ with mean $\boldsymbol{\mu}_k$ and precision $\boldsymbol{\Lambda}_k$. Then the probability density of the whole MAS consisting of K clusters is a Gaussian mixture of densities

$$p(\boldsymbol{x}_n) = \sum\nolimits_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right), \tag{5}$$

where π_k describe the *mixing coefficients*, with $0 \leq \pi_k \leq 1$ and $\sum_{k=1}^{K} \pi_k = 1$. The exact membership of an agent n to cluster k may be expressed by a K-dimensional binary vector-valued stochastic variable $\boldsymbol{z}_n$, for which $z_{nk} = 1$ and $z_{nl} = 0$, $\forall l \neq k$ holds. The prior probability density over z_{nk} is determined by

$$p(z_{nk} = 1) = \pi_k, \tag{6}$$

so that the marginal probability density for $\boldsymbol{z}$ may be expressed as

$$p(\boldsymbol{z}_n) = \prod_{k=1}^{K} \pi_k^{z_{nk}}. \tag{7}$$

The conditional probability over all memberships $\boldsymbol{Z} = \{\boldsymbol{z}_n\}$ given the mixing coefficients is therefore

$$p(\boldsymbol{Z} \mid \boldsymbol{\pi}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}}. \tag{8}$$

If the memberships $\boldsymbol{Z}$ as well as $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ are known, the conditional probability for the agents' positions is $\boldsymbol{X} = \{\boldsymbol{x}_n\}$

$$p(\boldsymbol{X} \mid \boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \mathcal{N}\left(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)^{z_{nk}}. \tag{9}$$

In this probability distribution, the agents' positions $\boldsymbol{x}_n$ are given and the unknown variables $\boldsymbol{\pi} = \{\pi_k\}$, $\boldsymbol{\mu} = \{\boldsymbol{\mu}_k\}$, $\boldsymbol{\Lambda} = \{\boldsymbol{\Lambda}_k\}$, $\boldsymbol{Z} = \{\boldsymbol{z}_n\}$ and the cluster number K have to be determined such that the positions $\boldsymbol{X}$ are explained by (5) as well as possible. By interpreting $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{Z}$ as stochastic variables, this corresponds to a maximization of the joint probability $p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$, which may be factored into conditional and prior probabilities according to

$$p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = p(\boldsymbol{X} \mid \boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) p(\boldsymbol{Z} \mid \boldsymbol{\pi}) p(\boldsymbol{\pi}) p(\boldsymbol{\mu} \mid \boldsymbol{\Lambda}) p(\boldsymbol{\Lambda}). \tag{10}$$

In order to determine the maximum posterior probability density $p(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda} \mid \boldsymbol{X})$ and the marginalized probability density $p(\boldsymbol{X})$, a variational approach is presented in Bishop (2006), which links their determination to the calculation of a functional $q(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$. For this, $\ln p(\boldsymbol{X}) = \mathcal{L}(q) + \mathrm{KL}(q \| p)$ with

$$\mathcal{L}(q) = \int q(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \ln \left[\frac{p(\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})}{q(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})}\right] d\boldsymbol{Z} d\boldsymbol{\pi} d\boldsymbol{\mu} d\boldsymbol{\Lambda}, \tag{11a}$$

$$\mathrm{KL}(q \| p) = -\int q(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \ln \left[\frac{p(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda} \mid \boldsymbol{x})}{q(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})}\right] d\boldsymbol{Z} d\boldsymbol{\pi} d\boldsymbol{\mu} d\boldsymbol{\Lambda}, \tag{11b}$$

is introduced, where $\mathcal{L}(q)$ is a lower bound on $\ln p(\boldsymbol{X})$ and KL is the *Kullback-Leibler-divergence* between the distributions q and p. A maximization of $\ln p(\boldsymbol{X})$ may now be achieved by maximizing the lower bound $\mathcal{L}(q)$ and by minimizing the Kullback-Leibler-divergence, which are both dependent on the newly introduced probability density q. In order to limit the search area for the functional, a factorization

$$q(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = q(\boldsymbol{Z})q(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \tag{12}$$

is being introduced in order to obtain a tractable solution of (11). Furthermore, some assumptions considering the prior probability densities $\boldsymbol{\pi}$, $\boldsymbol{\mu}$ und $\boldsymbol{\Lambda}$ have to be made. Because of (8), $p(\boldsymbol{Z}|\boldsymbol{\pi})$ has the form of a multinomial distribution. Making use of the concept of *conjugacy* a Dirichlet prior

$$p(\boldsymbol{\pi}) = \mathrm{Dir}\left(\boldsymbol{\pi} \mid \boldsymbol{\alpha}_0\right) = C(\boldsymbol{\alpha}_0) \cdot \prod_{k=1}^{K} \pi_k^{(\alpha_{0,k}-1)} \tag{13}$$

is assumed as prior distribution for $p(\boldsymbol{\pi})$, where $\boldsymbol{\alpha}$ denotes a vector of hyperparameters and $C(\cdot)$ is a constant. Likewise, because of the normal distribution of the agents' positions $\boldsymbol{x}_n$ within a cluster k, the conjugate prior of mean $\boldsymbol{\mu}_k$ given the precision $\boldsymbol{\Lambda}_k$ is again normally distributed, whereas the prior for $\boldsymbol{\Lambda}_k$ is a Wishart distribution. Thus, the joint probability density over $\boldsymbol{\mu}$ und $\boldsymbol{\Lambda}$ now takes the form

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = p(\boldsymbol{\mu}|\boldsymbol{\Lambda})p(\boldsymbol{\Lambda}) = \prod_{k=1}^{K} \mathcal{N}\left(\boldsymbol{\mu}_k \mid \boldsymbol{m}_0, (\beta_0\boldsymbol{\Lambda}_k)^{-1}\right) \mathcal{W}\left(\boldsymbol{\Lambda}_k \mid \boldsymbol{W}_0, \nu_0\right), \tag{14}$$

with parameter β_0 and ν_0 denoting the degrees of freedom of the Wishart distribution. By means of the variational approach, the optimal distributions $q^\star$ may be calculated in general. This leads to

$$q^\star(\boldsymbol{Z}) = \prod_{n=1}^{N}\prod_{k=1}^{K} r_{nk}^{z_{nk}}, \tag{15a}$$

$$r_{nk} = \frac{\rho_{nk}}{\sum_{j=1}^{K}\rho_{nj}}, \tag{15b}$$

$$\begin{aligned}\ln\rho_{nk} &= \mathbb{E}\left(\ln\pi_k\right) + \frac{1}{2}\mathbb{E}\left(\ln|\boldsymbol{\Lambda}_k|\right) - \frac{D}{2}\ln 2\pi \\ &\quad - \frac{1}{2}\mathbb{E}_{\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k}\left((\boldsymbol{x}_n - \overline{\boldsymbol{x}}_k)^\top \boldsymbol{\Lambda}_k(\boldsymbol{x}_n - \overline{\boldsymbol{x}}_k)\right),\end{aligned} \tag{15c}$$

as marginalized distribution over $\boldsymbol{Z}$, where the auxiliary variables

$$N_k = \sum_{n=1}^{N} r_{nk}, \tag{16a}$$

$$\overline{\boldsymbol{x}}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}\boldsymbol{x}_n, \tag{16b}$$

$$\boldsymbol{S}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}(\boldsymbol{x}_n - \overline{\boldsymbol{x}}_k)(\boldsymbol{x}_n - \overline{\boldsymbol{x}}_k)^\top \tag{16c}$$

have been introduced. The optimal prior for the mixing coefficient is then a Dirichlet-distribution

$$q^\star(\boldsymbol{\pi}) = \mathrm{Dir}\left(\boldsymbol{\pi} \mid \boldsymbol{\alpha}\right) \tag{17}$$

with hyperparameter components $\alpha_k = \alpha_0 + N_k$.

The expectation values of the mixing coefficients are thus given by $\mathbb{E}\left(\pi_k\right) = \frac{\alpha_{k'}}{\sum_{k'=1}^{K}\alpha_{k'}}$. Finally, the optimal prior distribution over $\boldsymbol{\mu}_k$ and $\boldsymbol{\Lambda}_k$ is calculated as

$$q^\star(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) = \mathcal{N}\left(\boldsymbol{\mu}_k \mid \boldsymbol{m}_k, (\beta_k\boldsymbol{\Lambda}_k)^{-1}\right) \mathcal{W}\left(\boldsymbol{\Lambda}_k \mid \boldsymbol{W}_k, \nu_k\right), \tag{18a}$$

with

$$\beta_k = \beta_0 + N_k, \tag{18b}$$

$$\boldsymbol{m}_k = \frac{1}{\beta_k}(\beta_0\boldsymbol{m}_0 + N_k\overline{\boldsymbol{x}}_k), \tag{18c}$$

$$\begin{aligned}\boldsymbol{W}_k^{-1} &= \boldsymbol{W}_0^{-1} + N_k\boldsymbol{S}_k \\ &\quad + \frac{\beta_0 N_k}{\beta_0 + N_k}(\overline{\boldsymbol{x}}_k - \boldsymbol{m}_0)(\overline{\boldsymbol{x}}_k - \boldsymbol{m}_0)^\top,\end{aligned} \tag{18d}$$

$$\nu_k = \nu_0 + N_k + 1. \tag{18e}$$

Their expected values are determined by $\mathbb{E}\left(\boldsymbol{\mu}_k\right) = \boldsymbol{m}_k$ and $\mathbb{E}\left(\boldsymbol{\Lambda}_k\right) = \nu_k \cdot \boldsymbol{W}_k$. Because of

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k}\left((\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{W}_k(\boldsymbol{x}_n - \boldsymbol{\mu}_k)\right) = \\ &\qquad D\beta_k^{-1} + \nu_k(\boldsymbol{x}_n - \boldsymbol{m}_k)^\top \boldsymbol{W}_k(\boldsymbol{x}_n - \boldsymbol{m}_k),\end{aligned} \tag{19}$$

the expected values for the memberships $\mathbb{E}\left(z_{nk}\right) = r_{nk}$ take the form

$$\begin{aligned}r_{nk} &\propto \tilde{\pi}_k\tilde{\Lambda}_k^{1/2} \\ &\cdot \exp\left\{-\frac{D}{2\beta_k} - \frac{\nu_k}{2}(\boldsymbol{x}_n - \boldsymbol{m}_k)^\top \boldsymbol{W}_k(\boldsymbol{x}_n - \boldsymbol{m}_k)\right\}\end{aligned} \tag{20a}$$

with

$$\begin{aligned}\ln\tilde{\Lambda}_k &:= \mathbb{E}\left(\ln|\boldsymbol{\Lambda}_k|\right) \\ &= \sum_{i=1}^{D}\psi\left(\frac{\nu_k + 1 - i}{2}\right) + D\ln 2 + \ln|\boldsymbol{W}_k|,\end{aligned} \tag{20b}$$

$$\ln\tilde{\pi}_k := \mathbb{E}\left(\ln\pi_k\right) = \psi\left(\alpha_k\right) - \psi\left(\sum\nolimits_{k'}\alpha_{k'}\right). \tag{20c}$$

Herein, $\psi(x) = \frac{d}{dx}\ln\Gamma(x)$ describes the *digamma function*. By means of these equations, the optimal model parameters of the Gaussian mixture density in (5) as well as $\boldsymbol{z}_n$ may be iteratively calculated with the following algorithm:

Algorithm 1. (Variational Bayes).

(1) Choice of an upper bound on the cluster number $K_{\max}$ and initialization of α_0, β_0, $\boldsymbol{m}_0$, $\boldsymbol{W}_0$, ν_0,
(2) *Variational Expectation*: Calculation of $\mathbb{E}\left(z_{nk}\right) = r_{nk}$ with (20) for fixed $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$,
(3) *Variational Maximization*: Calculation of optimal prior densities $q^\star(\boldsymbol{\pi})$ and $q^\star(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$ with (17) and (18) for fixed r_{nk},
(4) if convergence of probability densities $q(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ or the variables is not reached, go to step 2, else stop.

The advantage of this *Variational Bayes* is the ability of a soft partitioning of the agents' positions as well as an automatic determination of the model complexity, because only an upper bound $K_{\max}$ on the cluster number $K \leq K_{\max}$ has to be set in advance. Depending on the choice of $\boldsymbol{\alpha}$, more or less mixing coefficients $\pi_k \neq 0$ remain after convergence of the algorithm.

3.2 Distributed Variational Bayes

Considering a decentral implementation in an MAS, the aforementioned Alg. 1 allows a local calculation of the r_{nk} for fixed $\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ by each of the agents. On the other hand, it has the disadvantage that in order to compute the model parameters $\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ for fixed r_{nk}, global computations are necessary due to (16), which prevent a fully distributed implementation. With a consensus protocol like in (2), on the other hand, a decentral computation of the mean is possible. Therefore a strategy for the distributed calculation of the Variational Bayes Alg. 1 is the transformation of global computations into computations of mean values. For this to be achieved, (16) is first transformed into

$$N_k = \sum_{n=1}^{N} r_{nk} = N \cdot \omega_k, \tag{21a}$$

$$\overline{\boldsymbol{x}}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}\boldsymbol{x}_n = \frac{1}{N_k}\sum_{n=1}^{N} \boldsymbol{\tau}_{nk} = \frac{\tilde{\boldsymbol{\tau}}_k}{N_k}, \tag{21b}$$

$$\begin{aligned}\boldsymbol{S}_k &= \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}(\boldsymbol{x}_n - \overline{\boldsymbol{x}}_k)(\boldsymbol{x}_n - \overline{\boldsymbol{x}}_k)^\top \\ &= \frac{1}{N_k}\sum_{n=1}^{N} \boldsymbol{\rho}_{nk} = \frac{\tilde{\boldsymbol{\rho}}_k}{N_k}.\end{aligned} \tag{21c}$$

With help of the auxiliary variables

$$\omega_k = \frac{N_k}{N} = \frac{1}{N}\sum_{n=1}^{N} r_{nk}, \tag{22a}$$

$$\boldsymbol{\tau}_k = \frac{\tilde{\boldsymbol{\tau}}_k}{N} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\tau}_{nk}, \tag{22b}$$

$$\boldsymbol{\rho}_k = \frac{\tilde{\boldsymbol{\rho}}_k}{N} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\rho}_{nk} \tag{22c}$$

a calculation of (21) is possible with $N_k = N \cdot \omega_k$, $\overline{\boldsymbol{x}}_k = \frac{\boldsymbol{\tau}_k}{\omega_k}$, $\boldsymbol{S}_k = \frac{\boldsymbol{\rho}_k}{\omega_k}$. The advantage of the auxiliary variables in (22) is their affinity to mean value computations for r_{nk}, $\boldsymbol{\tau}_{nk}$ and $\boldsymbol{\rho}_{nk}$ of each agent. These may be carried out in a distributed manner by applying a consensus protocol for each of the variables in (22), e.g. with (2). For recovering (21), the number of agents N has to be known by all agents. This may also be easily computed in a distributed manner. With help of the auxiliary variables it is now possible to calculate the the prior probability densities $q^\star(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$. Thus, a fully distributed Variational Bayes is reached. Although being analytically slightly more complex than other clustering methods, the Variational Bayes approach allows for a complete description of the MAS formation as a mixture density, for which only an upper bound on the cluster number has to be set in advance. In addition to these benefits, this fully distributed version of the Variational Bayes may be applied to MAS with arbitrary connected communication graphs. Thus, the communication structure has not to be known to the agents and yet after convergence each agent has an estimate of the agents spatial distribution. This forms the basis of the N-consensus protocol.

4. THE N-CONSENSUS PROTOCOL

In order to reach an N-consensus, the agents have to be controlled by a signal $\boldsymbol{w}_n$ as in (4) in order to move to the center of clusters they are belonging to. Like the cluster centers $\boldsymbol{\mu}_k$ and memberships z_{nk}, the reference signal may as well be interpreted as a stochastic variable and its probability density can be denoted as

$$p(\boldsymbol{w}_n|\boldsymbol{r}_n, \boldsymbol{m}_n) = \mathcal{N}\left(\boldsymbol{w}_n \,\middle|\, \sum_k r_{nk}\boldsymbol{m}_{nk}, \boldsymbol{\Sigma}_w\right), \tag{23}$$

with $\mathbb{E}(\boldsymbol{z}_n) = \boldsymbol{r}_n$ and $\mathbb{E}(\boldsymbol{\mu}_n) = \boldsymbol{m}_n$. The complete algorithm for the calculation of the N-consensus is summarized in the following steps:

Algorithm 2. (N-consensus protocol).

(1) Choice of an upper bound of the cluster number $K_{\max}$ and initialization of parameters α_0, β_0, $\boldsymbol{m}_0$, $\boldsymbol{W}_0$, ν_0,$\boldsymbol{\Sigma}_w$,
(2) *Variational Expectation*: Calculation of r_{nk} with (20),
(3) *Consensus*: Distributed calculation of ω_k, $\boldsymbol{\tau}_k$ and $\boldsymbol{\rho}_k$ in (22) with a consensus protocol,
(4) *Variational Maximization*: Computation of intermediate variables N_k, $\overline{\boldsymbol{x}}_k$, $\boldsymbol{S}_k$ with (21), calculation of optimal prior probability densities $q^\star(\boldsymbol{\pi})$ and $q^\star(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$ with (17) and (18) for fixed r_{nk},
(5) local computation of new reference signals $\boldsymbol{w}_n$ with (23),
(6) if convergence of probability densities $q(\boldsymbol{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ or variables is not reached, go to step 2, else stop.

In Alg. 2, a distributed clustering is reached by means of a Variational Bayes working on the initial agents' positions $\boldsymbol{x}_n(0)$. Depending on the estimated mixture density representing the whole MAS, each agent derives a control signal. Considering the overall system dynamics, this process may be seen as an open-loop control involving the distributed Variational Bayes and the MAS. It is now tempting to ask whether a consideration of the current agents' positions $\boldsymbol{x}_n(t)$ in the clustering process may lead to an improvement of the system dynamics. This equivalence to a feedback control is shown in Fig. 2. Because the agents are solely moving towards the centers of their clusters, such an adaption should not change the result of the final N-consensus, although the estimated mixture density will be different. The advantage of this adaption will be a faster convergence of the whole process due to sharper cluster shaping over time.

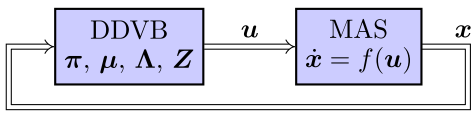

Fig. 2. Feedback control of MAS with distributed Dynamical Variational Bayes (DDVB)

Considering an implementation of such a *distributed Dynamical Variational Bayes (DDVB)*, one has to take care of r_{nk}, which converge asymptotically to 0 or 1. For agents adapting their positions $\boldsymbol{x}_n$, this might cause problems in cases where the agents follow their control input faster than r_{nk} are converging towards a fixed value $r_{nk} \in \{0, 1\}$. This would lead to a convergence of all agents' positions onto one single synchronous state, independent of their original formation. By carrying out a hard decision according to

$$\tilde{r}_{nk} = \begin{cases} 1, & \text{if } \exists\, r_{nl} : r_{nl} > \tau_r \wedge k = l \\ 0, & \text{if } \exists\, r_{nl} : r_{nl} > \tau_r \wedge k \neq l \\ r_{nk}, & \text{otherwise} \end{cases}, \tag{24}$$

this undesirable behavior may be avoided. The threshold τ_r may be chosen rather restrictive, e.g. $\tau_r \geq 0.95$. Thus, an agent is assumed to fully belong to a certain cluster whenever its expected membership exceeds τ_r. Another problem results from the expected value $\mathbb{E}(\boldsymbol{\mu}_k) = \boldsymbol{m}_k$ in

(18c). In order for the Variational Bayes not to run into singularities for $D \geq 2$ and data points collapsing onto a cluster center, $\beta_0 \neq 0$ has to be chosen. In case of an adaption of $\boldsymbol{x}_n$ and $\boldsymbol{m}_0 = \boldsymbol{0}$, this offset causes a drift of the agents towards the origin of the coordinate system due to

$$\boldsymbol{m}_k = \frac{\beta_0 \boldsymbol{m}_0}{\beta_k} + \frac{N_k \overline{\boldsymbol{x}}_k}{\beta_k} = \frac{1}{1 + \frac{\beta_0}{N_k}} \overline{\boldsymbol{x}}_k, \tag{25}$$

which forms a sequence converging to zero. This is avoided by a correction of the Variational Bayes with

$$\tilde{\boldsymbol{m}}_k = \boldsymbol{m}_k + \frac{\beta_0}{\beta_k} \cdot \overline{\boldsymbol{x}}_k = \frac{N_k}{N_k + \beta_0} \overline{\boldsymbol{x}}_k + \frac{\beta_0}{\beta_k} \cdot \overline{\boldsymbol{x}}_k = \overline{\boldsymbol{x}}_k. \tag{26}$$

Algorithm 3. (Dynamical N-consensus protocol). Akin to Alg. 2, but with adaption of agents' positions $\boldsymbol{x}_n(t)$ within the DVB and $\boldsymbol{w}_n(t) = \sum_k \tilde{r}_{nk}(t) \cdot \tilde{\boldsymbol{m}}_k(t)$.

This differs from Alg. 2, because the Variational Bayes now uses the current agents' positions $\boldsymbol{x}_n(t)$ instead of the inital values $\boldsymbol{x}_n(0)$.

5. THE N-CONSENSUS MODEL

In Bishop (2006) it is shown that probabilistic dependencies may be visualized by graphical models. As an example, the joint probability in (10) may be represented by the Bayesian Network in Fig. 3a. Therein, the observed stochastic variables $\boldsymbol{x}_n$ are being represented by filled circles, whereas unobserved (latent) variables are depicted by empty circles. The rounded box around the variables $\boldsymbol{z}_n$ und $\boldsymbol{x}_n$ emphasizes that they are given for each of the N agents, whereas the parameters $\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ of the mixture density appear only once.

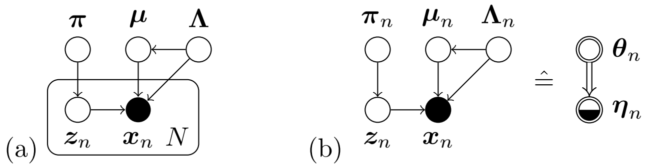

Fig. 3. (a) Mixture of Gaussians as Bayesian Network and (b) equivalent compact notation using vector valued stochastic variables

Because in the distributed Variational Bayes each agent calculates an identical copy of the model parameters $\boldsymbol{\pi}_n$, $\boldsymbol{\mu}_n$ and $\boldsymbol{\Lambda}_n$ using the consensus protocol, this can be taken into consideration graphically by Fig. 3b, which also shows a compact notation using *hypernodes*. Therein, the vector valued node $\boldsymbol{\eta}_n$ combines the local variables $\boldsymbol{x}_n$ and $\boldsymbol{z}_n$ of each agent and $\boldsymbol{\theta}_n$ represents the set of model parameters $\boldsymbol{\pi}_n$, $\boldsymbol{\mu}_n$ and $\boldsymbol{\Lambda}_n$ that were calculated by the n-th agent.

The probabilistic dependencies between the reference signal $\boldsymbol{w}_n$ and the hypernodes $\boldsymbol{\eta}_n$ and $\boldsymbol{\theta}_n$ may equivalently be visualized as shown in Fig. 4.

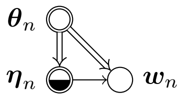

Fig. 4. Dependency of reference signal from vectorial nodes $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$

The partial graphical models may be combined to a unifying graph, which is illustrated for the MAS example in Fig. 1. Each agent n possesses a local variable $\boldsymbol{\eta}_n$ and a global variable $\boldsymbol{\theta}_n$ as well as a reference signal $\boldsymbol{w}_n$, therefore being described by three nodes as shown in Fig. 4. In order to compute (22) in a distributed manner, the agents have to transfer information over the communication graph, which is symbolized by the vectorial connections between global variables $\boldsymbol{\theta}_n$. This results in the *N-consensus model* depicted in Fig. 5, which shows probabilistic dependencies as well as the communication structure.

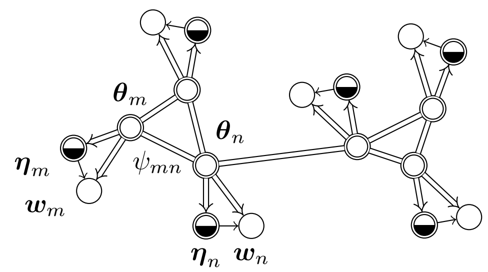

Fig. 5. N-consensus model

When introducing Gaussian potentials

$$\psi_{mn}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_n) = e^{-(\boldsymbol{\theta}_m - \boldsymbol{\theta}_n)^2} \tag{27}$$

for the edges of each two hypernodes $\boldsymbol{\theta}_m$, $\boldsymbol{\theta}_n$ with $m \neq n$ in order to characterize the probabilistic dependencies among the global variables of the agents, the N-consensus model describes a potential

$$\Xi = \prod_n p(\boldsymbol{\eta}_n, \boldsymbol{\theta}_n) \cdot \prod_{\{m,n\}} \psi_{mn}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_n) \cdot \prod_n p(\boldsymbol{w}_n \,|\, \boldsymbol{\eta}_n, \boldsymbol{\theta}_n), \tag{28}$$

which includes the joint probabilities $p(\boldsymbol{\eta}_n, \boldsymbol{\theta}_n)$ being calculated by the agents with (10) as well as the reference signals $p(\boldsymbol{w}_n \,|\, \boldsymbol{\eta}_n, \boldsymbol{\theta}_n)$ derived from the model estimates. The synchronization among the global variables $\boldsymbol{\theta}_m$, $\boldsymbol{\theta}_n$ is taken into account by $\psi_{mn}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_n)$. Altogether, an N-consensus of the agents is reached by maximizing (28).

If a dynamical Variational Bayes is used in order to improve the overall system dynamics, the complete system in Fig. 2 may also be visualized as a Dynamical Bayesian Network (DBN) as shown in Fig. 6. Herein, the control signals are being used in the following time steps as new position values that are to be clustered. In addition, the agents are measuring new position values $\boldsymbol{y}_n(t+1)$ that are identical with $\boldsymbol{x}_n(t+1)$ and thus with $\boldsymbol{w}_n(t)$ due to the integrator dynamics of the agents. When combining the latent stochastic variables in a hypernode $\boldsymbol{\Omega} = \{\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}\}$, the DBN reveals an analogy to a Markov chain of order one, which is shown in Fig. 7.

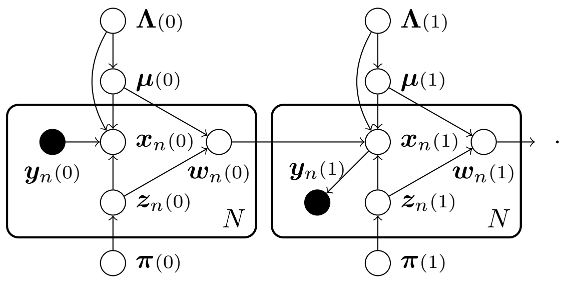

Fig. 6. N-consensus as DBN unrolled over two time steps

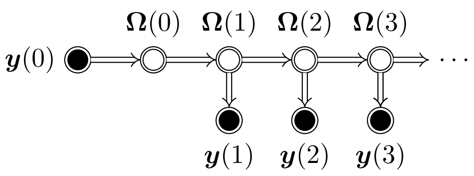

Fig. 7. N-consensus representation as Markov chain

6. SIMULATION RESULTS

For the MAS in Fig. 1, the agent trajectories until convergence upon N-consensus are shown in Fig. 8. In the same figure, a contour plot of the estimated mixture density in case of a Variational Bayes without dynamical adaption of the position values is drawn as well. A second example with an MAS consisting of 100 agents, Fig. 9 shows the result of an N-consensus with dynamical adaption of the Variational Bayes. For this example, the summed difference between the agents current position and their final position

$$F(t) = \sum_{n=1}^{N} \left\| \sum_{k=1}^{K} \tilde{r}_{nk}(t) \cdot \tilde{\boldsymbol{m}}_k^{(n)}(t) - \tilde{r}_{nk}(T_{\text{end}}) \cdot \tilde{\boldsymbol{m}}_k^{(n)}(T_{\text{end}}) \right\| \quad (29)$$

is shown in Fig. 10, which can be interpreted as a *control error*. As can be seen, the N-consensus converges faster in case of a dynamical adaption of position values used within the Variational Bayes compared to a clustering which is only executed on the initial agents' positions.

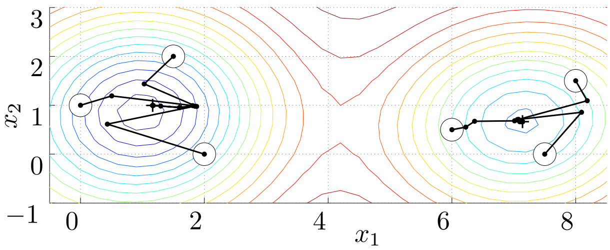

Fig. 8. Agent trajectories until reach of N-consensus

7. CONCLUSION

The N-consensus was addressed for the first time, describing a combined clustering and synchronization upon multiple states within an MAS. Based on existing probabilistic methods, it was shown that a Variational Bayes in MAS with arbitrary connected topology may be computed in a distributed manner. In order to visualize probabilistic dependencies, a graphical model was presented which describes a potential that is maximized when a N-consensus is reached. By dynamically adapting the position values used for the clustering process, an improvement in the overal system dynamics was gained. At the same time, the analogy of this process to a first order Markov chain was shown. In order to synchronize auxiliary variables, a consensus filter was utilized.

Further research is necessary on how to include the actual dynamics of the agents into the framework as well as their impact on the system dynamics and the N-consensus. Also, the effect of weighted communication graphs and switching topologies remains an open question. The framework presented might be an even more general solution

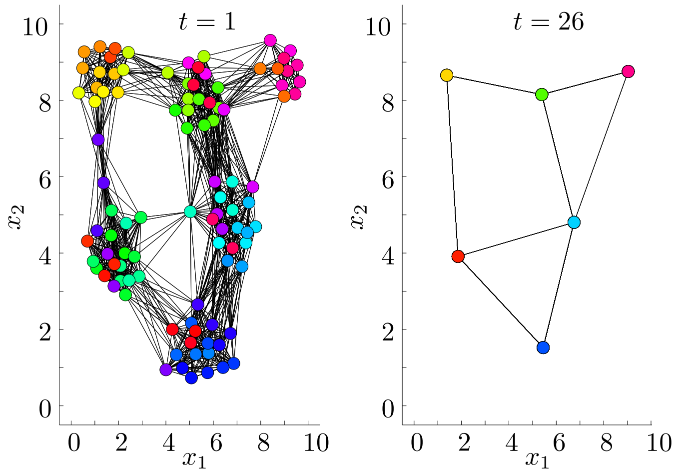

Fig. 9. N-Consensus in MAS with 100 agents

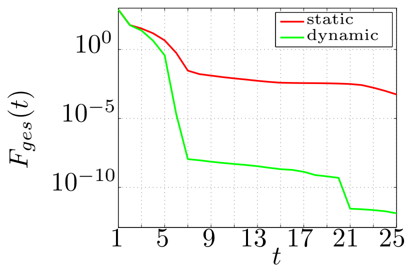

Fig. 10. Control error over time for MAS with 100 agents

method for problems involving distributed clustering and synchronization.

REFERENCES

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Forero, P.A., Cano, A., and Giannakis, G.B. (2008). Consensus-based K-means Algorithm for Distributed Learning using Wireless Sensor Networks. In *Workshop on Sensors, Signal and Information Processing*.

Frey, B.J. and Dueck, D. (2007). Clustering by Passing Messages Between Data Points. *Science*, 315, 972–976.

Gu, D. (2008). Distributed EM Algorithm for Gaussian Mixtures in Sensor Networks. *IEEE Transactions on Neural Networks*, 19(7), 1154–1166.

Listmann, K., Wahrburg, A., Strubel, J., Adamy, J., and Konigorski, U. (2011). Partial-state synchronization of linear heterogeneous multi-agent systems. In *50th IEEE CDC and ECC, 2011*, 3440–3445.

Moallemi, C.C. and Van Roy, B. (2006). Consensus Propagation. *IEEE Transactions on Information Theory*, 52(11), 4753–4766.

Saber, R. and Murray, R. (2003). Consensus protocols for networks of dynamic agents. In *American Control Conference, 2003. Proceedings of the 2003*, volume 2, 951 – 956.

Safarinejadian, B., Menhaj, M., and Karrari, M. (2010). Distributed variational Bayesian algorithms for Gaussian mixtures in sensor networks. *Signal Processing*, 90(4), 1197–1208.

Yu, J. and Wang, L. (2009). Group consensus of multi-agent systems with undirected communication graphs. In *Asian Control Conference, 2009. ASCC 2009. 7th*, 105 –110.