Automatica 48 (2012) 374–380

Contents lists available at SciVerse ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

A hyperparameter consensus method for agreement under uncertainty☆

Cameron S.R. Fraser[a], Luca F. Bertuccelli[c], Han-Lim Choi[b,1], Jonathan P. How[a]

[a] 77 Massachusetts Ave., Cambridge, MA 02139, USA
[b] 291 Daehak-ro, Yuseong, Deajeon 305-701, Republic of Korea
[c] 411 Silver Lane, East Hartford, CT 06108, USA

ARTICLE INFO

Article history:
Received 5 July 2010
Received in revised form
13 February 2011
Accepted 19 July 2011
Available online 28 December 2011

Keywords:
Consensus
Bayesian parameter estimation
Non-Gaussian uncertainty
Hyperparameter
Conjugacy

ABSTRACT

This paper addresses the problem of information consensus in a team of networked agents by presenting a generic consensus method that permits agreement to a Bayesian fusion of uncertain local parameter estimates. In particular, the method utilizes the concept of *conjugacy* of probability distributions to achieve a steady-state estimate consistent with a Bayesian combination of each agent's local knowledge, without requiring complex channel filters or being limited to normally distributed uncertainties. It is shown that this algorithm, termed *hyperparameter consensus*, is adaptable to many local uncertainty distributions within the exponential family, and will converge to a Bayesian fusion of local estimates with some standard assumptions on the network topology.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Future unmanned systems missions will require the interaction of distributed and networked groups of agents that can sense the environment and broadcast the acquired information to the other members of the team. Consensus algorithms operating over communication networks have found many applications in distributed decision making from flocking to rendezvous (Beard & Stepanyan, 2003; Fax & Murray, 2004; Jadbabaie, Lin, & Morse, 2003; Lin, Morse, & Anderson, 2003; Olfati-Saber, Fax, and Murray, 2007; Ren, 2006; Ren, Beard, & Atkins, 2007; Ren, Beard, & McLain, 2005), and guarantee convergence of the fleet's situational awareness while remaining computationally inexpensive even over large, complex networks, and many different dynamic network topologies (Cortés, 2008; Hatano & Mesbahi, 2005; Tahbaz-Salehi & Jadbabaie, 2008; Wu, 2006; Zhu & Martínez, 2010). However, most of these approaches (e.g., Fax & Murray, 2004, Jadbabaie et al., 2003 and Lin et al., 2003) have assumed that (a) there is no uncertainty in each agent's local estimate of the quantity of interest, and (b) there is no *true* value to which the agents are attempting to agree (i.e., the agents having agreed is the only necessary result).

Kalman consensus approaches were devised to allow for local uncertainties that can be modeled using a Gaussian distribution (Alighanbari & How, 2008; Ren, Beard, & Kingston, 2005). The resulting consensus is influenced more heavily by agents with smaller covariance (therefore higher certainty) in their estimates. Unfortunately, applying Kalman filter-based consensus methods to the mean and covariance of non-Gaussian distributions may produce a steady-state estimate that is biased away from the Bayesian estimate obtained by using the true form of the distribution (Bertuccelli, How, 2009).

In situations where there is some *true* value to agree upon or learn, Bayesian decentralized data and sensor fusion methods are able to determine the best combined Bayesian parameter estimate given a set of objective observations (Grime & Durrant-Whyte, 1994; Grime, Durrant-Whyte, & Ho, 1992; Makarenko & Durrant-Whyte, 2006; Waltz & Llinas, 1990). Unfortunately, as noted in Grime and Durrant-Whyte (1994) and Grime et al. (1992) these approaches require channel filters to handle common or repeated information in messages between neighboring nodes, meaning that even for known network topologies other than the simplest fully connected and tree networks there is no simple channel filter algorithm available. Some recent approaches have combined the Kalman consensus and data fusion methods to achieve scalable, representative information fusion results without

☆ The material in this paper was partially presented and preliminarily presented at the AIAA Guidance, Navigation, and Control Conference, August 10–13, 2009, Chicago, Illinois, USA, and at the American Control Conference, June 30–July 2, 2010, Baltimore, Maryland, USA. This paper was recommended for publication in revised form by Associate Editor Giancarlo Ferrari-Trecate under the direction of Editor Ian R. Petersen.

E-mail addresses: csrfraserc@gmail.com (C.S.R. Fraser), bertuclf@utrc.utc.com (L.F. Bertuccelli), hanlimc@kaist.ac.kr (H.-L. Choi), jhow@mit.edu (J.P. How).

1 Tel.: +82 42 350 3727; fax: +82 42 350 3727.

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.11.003

requiring complex channel filters or specific network topologies (Olfati-Saber, 2006; Olfati-Saber & Shamma, 2005; Xiao, Boyd, & Lall, 2005). In particular, Olfati-Saber and Shamma (2005) and Olfati-Saber (2006) utilized dynamic-average consensus filters to achieve an approximate distributed Kalman filter, while Xiao et al. (2005) implemented a linear consensus protocol on the parameters of the information form of the Kalman filter, permitting the agents using either method to execute a Bayesian fusion of normally-distributed random variables. However, since the methods in Olfati-Saber (2006), Olfati-Saber and Shamma (2005) and Xiao et al. (2005) are derived specifically for normally-distributed uncertainties, they can produce biased results if the local distributions are non-Gaussian.

This article extends the current state-of-the-art and presents a novel consensus method to allow a group of networked agents to agree on the Bayesian fusion of their local uncertain parameter estimates under a range of non-Gaussian uncertainties without the need for complex channel filters. A primary contribution of this paper and a key to achieving this result is to exploit the *conjugacy* property of parameterized probability distributions (Gelman, Carlin, Stern, & Rubin, 2004) to shift the focus of consensus from the parameter of interest to the parameters describing the local uncertainty distributions, called hyperparameters. The resulting agreement on the hyperparameters leads inherently to agreement on the parameter, but intrinsically maintains the proper weighting due to each agent's uncertainty. This generic approach, termed *hyperparameter consensus* (HPC), is adaptable to many probability distributions in the exponential family and is proven to converge to a Bayesian fusion of local uncertain estimates over any strongly connected networks, assuming some level of local knowledge about the network topology.

2. Background

2.1. Bayesian parameter estimation

This section introduces some of the terminology that will be used in this paper (see Gelman et al., 2004 for more information).

$f_{X|Y}(X=x|Y=y)$ defines a generic probability distribution on a random variable (R.V.) X, conditioned on Y, and is assumed to be discrete or continuous as required by the allowed domain of X. For brevity, it will often be denoted simply as $f_{X|Y}(x|y)$ or $f_{X|Y}$.

Θ is the R.V. denoting an estimate of the parameter of interest that the agents are attempting to agree upon. Realizations of this R.V. are denoted by θ.

Z is the R.V. of which each measurement, z, is a realization. It is defined by a stationary stochastic process governed by the parameter of interest, $f_{Z|\Theta}(z|\theta)$.

Ω represents the set of *hyperparameters*, ω, a (possibly multivariate) real-valued parameterization of the initial, *prior* distribution on Θ before a measurement.

The Bayesian update of a prior distribution on θ, $f_{\Theta|\Omega}(\theta|\omega)$, for a realization, z, from a stochastic measurement (with measurement model $f_{Z|\Theta}(z|\theta)$), results in a posterior distribution found by:

$$f_{\Theta|Z,\Omega}(\theta|z,\omega)=\frac{f_{Z|\Theta}(z|\theta)f_{\Theta|\Omega}(\theta|\omega)}{\int f_{Z,\Theta|\Omega}(z,\theta|\omega)d\theta}, \tag{1}$$

where $f_{\Theta|\Omega}(\theta|\omega)$ represents the prior, $f_{Z|\Theta}(z|\theta)$ is also known as the likelihood function, and the denominator on the right-hand-side ensures the posterior integrates to unity.

Many traditional likelihoods have associated prior distributions called *conjugate distributions* for which the Bayesian update can be evaluated in closed form, thereby avoiding difficulties in the numerical integration of the denominator in (1).

Definition 1 (*Conjugate Distributions*)**.** For conjugate distributions, the posterior distribution is of the same form[2], but uses an updated set of hyperparameters: $f_{\Theta|Z,\Omega}(\theta|z,\omega)=f_{\Theta|\Omega}(\theta|\omega_{\text{posterior}})$.

In many cases, using a particular selection of hyperparameters yields an *additive update* of the form:

$$\omega_{\text{posterior}}=h(z,f_{Z|\Theta})+\omega_{\text{prior}}, \tag{2}$$

where z is realized from the measurement model, and $h(\cdot,\cdot)$ is a possibly nonlinear function of the observed measurement and the measurement model and results in a non-negative change in hyperparameters. One special case of particular interest is when the hyperparamter is updated simply by the observed outcome, z. In a simplified notation, this hyperparameter update becomes: $\omega\leftarrow z+\omega$, where $\leftarrow$ denotes the update step.

Assumption 1. This work assumes that the local distributions are conjugate priors for some likelihood function and that an additive hyperparameter update in the form of (2) exists for the local priors. This is henceforth called the *conjugacy property*.

Note that Assumption 1 also implies that the hyperparameters are exchangeable, meaning that measurements can be incorporated in any order, and that if given only the current hyperparameters, it is impossible to determine which hyperparameters were contributed by which measurement.

2.2. Linear consensus

This paper utilizes linear consensus results as outlined in Olfati-Saber et al. (2007), Ren et al. (2007), and the reader is referred there for more information. The linear consensus protocol operates over a network defined by a strongly connected *directed* graph $\mathcal{G}$, where each agent is represented by a node, there is a directed edge from agent i to j if and only if agent i can transmit to j, and each agent is assumed to be able to talk to itself. Let $\mathcal{N}\triangleq\{1,2,\ldots,N\}$ denote the set of all the nodes, and let $\mathcal{N}_i$ denote the set of agents that can transmit to i, and let the adjacency matrix, A, for the graph be defined by entries $a_{ij}\geq 0$, which may be non-zero if there exists an edge from j to i and are zero otherwise.

The consensus protocol on a scalar ξ_i at iteration k is:

$$\xi_i[k+1]=\xi_i[k]+\epsilon\sum_{j\in\mathcal{N}_i}(\xi_j[k]-\xi_i[k])\triangleq\sum_{j\in\mathcal{N}}a_{ij}\xi_j[k]$$

$$\xi[k+1]=A\xi[k]=A^{k+1}\xi[0]. \tag{3}$$

For (3) to converge, the weighting constant, ϵ, is subject to $\epsilon\in(0,\ 1/\max_i|\mathcal{N}_i|)$ (Olfati-Saber et al., 2007). The result of this linear consensus protocol is found in the limit as

$$\lim_{k\to\infty}\xi_i[k+1]=\nu^T\xi[0]=\sum_{j\in\mathcal{N}}\nu_j\xi_j[0],$$

where ν is the *consensus eigenvector*, the normalized left-eigenvector of the adjacency matrix corresponding to its 1-eigenvalue (see Ren et al., 2007), which satisfies $\lim_{k\to\infty}A^k=\mathbf{1}\nu^T$ and $\nu^TA=\nu^T$, where $\mathbf{1}$ is an N-dimensional vector all of whose entries are unity. If the network is balanced, such that $\sum_{j=1}^N a_{ij}=\sum_{j=1}^N a_{ji}$, $\forall i\in\mathcal{N}$, then all $\nu_i=1/N$ and the resulting consensus value is an arithmetic average of the initial conditions. If the topology of an unbalanced network is known, then each agent can calculate ν and augment its initial condition to preserve the arithmetic average-consensus result. This is achieved by weighting the initial conditions as $\xi_i[0]=\xi_i[0^-]/(\nu_iN)$, where $\xi_i[0^-]$ denotes the un-weighted information for agent i (Ren et al., 2007).

[2] The terminology that one distribution is "of the same form" as another distribution will be understood to mean that the two distributions share the same functional form (e.g., both normally distributed) as the prior but may have different parameters (e.g., different mean and/or variance).

3. Hyperparameter consensus (HPC) method

3.1. Bayesian fused estimate

Consider N agents attempting to combine their uncertain estimates of Θ, where each agent maintains a local distribution on Θ of the same form but parameterized by uniquely valued local hyperparameters, ω_i for $i \in \mathcal{N}$. In this section, the fused Bayesian posterior distribution on Θ will be expressed as a function of local hyperparameters ω_i and the shared hyperparameters, denoted as ω^-, representing common information amongst the team.

3.1.1. Information structure

In this paper, globally shared information is allowed as long as the information is common to *all* agents. Since the hyperparameters are updated in a linear fashion by (2) and completely define each agent's knowledge of Θ through $f_{\Theta|\Omega}$, the shared information can be represented by a corresponding subset of each agent's hyperparameters. There are two representative sources of shared information: (a) a common initial prior distribution, $f_{\Theta|\Omega}(\theta|\omega_0)$, which is defined by the hyperparameter ω_0; and (b) a shared measurement among all agents. Considering the set of M_*^- measurements shared amongst the entire team, $Z_*^- \triangleq \{z_{*,1}^-, \ldots, z_{*,M_*^-}^-\}$, where the contribution of each of these shared measurements to the hyperparameters follows the additive update in (2), then the overall effect is represented by the addition of $\sum_{j=1}^{M_*^-} h(z_{*,j}^-, f_{Z|\Theta}) = \sum_{z\in Z_*^-} h(z, f_{Z|\Theta})$ to the hyperparameters. Therefore, the total shared information by the two aforementioned types can be described by the shared hyperparameter, ω^-:

$$\omega^- = \omega_0 + \sum_{z\in Z_*^-} h(z, f_{Z|\Theta}). \tag{4}$$

Similarly, there are two representative sources of information that are unique to a particular agent i: (a) a unique local prior defined by a corresponding hyperparameter, $\omega_{i,0}$; and (b) measurements that are uniquely available to agent i. Considering the set of M_i^- measurements unique to agent i, denoted as $Z_i^- \triangleq \{z_{i,1}^-, \ldots, z_{i,M_i}^-\}$, their contribution to agent i's local information is described by $\sum_{z\in Z_i^-} h(z, f_{Z|\Theta})$. Therefore, the local information unique to the agent is contained in a set of hyperparameters, denoted as $\Delta\omega_i$ that satisfies

$$\Delta\omega_i = \omega_{i,0} + \sum_{z\in Z_i^-} h(z, f_{Z|\Theta}). \tag{5}$$

The hyperparameters in (4) and (5) determine an agent's information structure before initiating consensus, such that the following summation is satisfied:

$$\omega_i = \omega^- + \Delta\omega_i, \quad \forall i \in \mathcal{N}. \tag{6}$$

Note that each agent's local estimate is independent from every other agent's local estimate when conditioned on the globally shared information, ω^-, or, equivalently, that the $\Delta\omega_i$ are all independent of each other. Since the common information must only be counted once when coming to agreement lest the result will be biased towards the repeated shared information, the following assumption is also made.

Assumption 2. It is assumed that all ω^- are known to be shared by all agents before and during a consensus, that any shared measurements are shared among *all* agents, and that all agents know if a given measurement is shared.

Finally, in addition to Assumptions 1 and 2, the following is required for the derivations in the next section.

Assumption 3. It is assumed that all measurements are modeled using a likelihood function for which the local distributions are conjugate and $h(z, f_{Z|\Theta})$ is well defined.

3.1.2. Derivation of the fused estimate

Lemma 1 (*Local Conjugacy*)**.** *For each agent's local distribution on Θ based on their local hyperparameter, i.e., $f_{\Theta|\Omega}(\theta|\omega_i)$, $\forall i \in \mathcal{N}$, the following properties hold:*

(a) *For any agent $i \in \mathcal{N}$, its local distribution on Θ can be determined as the product of a term representing the agent's unique local information, which will be called its pseudo-likelihood, and a prior distribution using the globally shared hyperparameters (for brevity, this will henceforth be called the global prior):*

$$\underbrace{f_{\Theta|\Omega}(\theta|\omega_i)}_{\text{local distribution}} = \underbrace{f_{\Delta\Omega|\Theta}(\Delta\omega_i|\theta)}_{\text{pseudo-likelihood}} \underbrace{f_{\Theta|\Omega}(\theta|\omega^-)}_{\text{global prior}} /\eta_i \tag{7}$$

where the normalizing factor η_i is defined as $\eta_i = \int f_{\Delta\Omega|\Theta}(\Delta\omega_i|\theta) f_{\Theta|\Omega}(\theta|\omega^-)d\theta$.

(b) *The global prior $f_{\Theta|\Omega}(\theta|\omega^-)$ is conjugate to the pseudo-likelihood $f_{\Delta\Omega|\Theta}(\Delta\omega_i|\theta)$, $\forall i \in \mathcal{N}$.*

Proof. (a) Since the information unique to each agent can be completely described by the unique local hyperparameters, $\Delta\omega_i$, each agent's local information $f_{\Theta|\Omega_i}$ can be decomposed into:

$$f_{\Theta|\Omega}(\theta|\omega_i) = f_{\Theta|\Omega}(\theta|\Delta\omega_i \cup \omega^-) = f_{\Theta|\Delta\Omega,\Omega}(\theta|\Delta\omega_i, \omega^-).$$

Applying Bayes' rule to the above yields

$$\begin{aligned} f_{\Theta|\Omega}(\theta|\omega_i) &= \frac{f_{\Delta\Omega|\Theta,\Omega}(\Delta\omega_i|\theta,\omega^-) f_{\Theta|\Omega}(\theta|\omega^-)}{f_{\Delta\Omega|\Omega}(\Delta\omega_i|\omega^-)} \\ &= \frac{f_{\Delta\Omega|\Theta}(\Delta\omega_i|\theta) f_{\Theta|\Omega}(\theta|\omega^-)}{\int f_{\Theta,\Delta\Omega|\Omega}(\theta, \Delta\omega_i|\omega^-)d\theta}, \\ &= \frac{f_{\Delta\Omega|\Theta}(\Delta\omega_i|\theta) f_{\Theta|\Omega}(\theta|\omega^-)}{\int f_{\Delta\Omega|\Theta}(\Delta\omega_i|\theta) f_{\Theta|\Omega}(\theta|\omega^-)d\theta}, \end{aligned}$$

where it has been used that $\Delta\omega_i$ is independent of ω^-.

(b) The conjugacy is straightforwardly inferred from the observation that both the prior and the posterior terms in (7) are of the same form. □

Lemma 2 (*Recursive Conjugacy*)**.** *Consider the posterior distribution on Θ after incorporating local information from agent 1 through n (or equivalently, any permutation of size n out of $\mathcal{N}$),*

$$f_{\Theta|\Omega_{1:n}}, \tag{8}$$

for some $n \geq 1$, where $\Omega_{1:n} \triangleq \{\Omega_1, \ldots, \Omega_n\}$ and $\omega_{1:n} \triangleq \{\omega_1, \ldots, \omega_n\}$. Then, the following properties hold:

(a) *$f_{\Theta|\Omega_{1:n}}(\theta|\omega_{1:n})$ can be expressed as a recursive form:*

$$\underbrace{f_{\Theta|\Omega_{1:n}}(\theta|\omega_{1:n})}_{\text{posterior at } n} = \underbrace{f_{\Delta\Omega|\Theta}(\Delta\omega_n|\theta)}_{\text{pseudo-likelihood}} \times \underbrace{f_{\Theta|\Omega_{1:n-1}}(\theta|\omega_{1:n-1})}_{\text{prior at } n\text{; posterior at } n-1} /\eta_{1:n}, \tag{9}$$

where $\eta_{1:n} = \int f_{\Delta\Omega|\Theta}(\Delta\omega_n|\theta) f_{\Theta|\Omega_{1:n-1}}(\theta|\omega_{1:n-1})d\theta$ and $\omega_{1:0} \triangleq \omega^-$.

(b) *$f_{\Theta|\Omega_{1:n}}(\theta|\omega_{1:n})$ is conjugate to the pseudo-likelihood function in (9), is of the same form as the global prior and, therefore, can be defined by a single hyperparameter $\overline{\omega}_n = \omega^- + \sum_{i=1}^n \Delta\omega_i$, such that*

$$f_{\Theta|\Omega_{1:n}}(\theta|\omega_{1:n}) = f_{\Theta|\Omega}(\theta|\overline{\omega}_n). \tag{10}$$

Proof. (a) The posterior distribution can be decomposed as $f_{\Theta|\Omega_{1:n}}(\theta|\omega_{1:n}) = f_{\Theta|\Delta\Omega_{1:n},\Omega}(\theta|\Delta\omega_{1:n}, \omega^-)$, for any $n \in \mathcal{N}$; applying Bayes' rule and utilizing the independence of $\Delta\omega_i$'s yields (9).

(b) The proof is by induction. First, for $n = 1$, $\omega_{1:n} = \omega_1$, and Lemma 1 with $i = 1$ states that the global prior is conjugate to the pseudo-likelihood $f_{\Delta\Omega|\Theta}(\Delta\omega_1|\theta)$. By the definition of $\Delta\omega_i$, the corresponding hyperparameter update resulting in the posterior $f_{\Theta|\Omega_{1:1}}(\theta|\omega_{1:1})$ is

$$\overline{\omega}_1 = \omega_1 = \Delta\omega_1 + \omega^-. \tag{11}$$

Second, suppose that the posterior distribution at $m-1$ (equivalent to the prior distribution at m) is of the same form as the global prior and defined by a hyperparameter $\overline{\omega}_{m-1} = \omega^- + \sum_{i=1}^{m-1} \Delta\omega_i$. From Lemma 1(b), the global prior is conjugate to the pseudo-likelihood $f_{\Delta\Omega|\Theta}(\Delta\omega_m|\theta)$, implying that the prior at m is as well. Therefore, the posterior at m can be determined through the additive update $\overline{\omega}_m = \Delta\omega_m + \overline{\omega}_{m-1} = \omega^- + \sum_{i=1}^{m} \Delta\omega_i$, which completes the induction proof. □

Theorem 1 (*Fused Distribution*). *Consider a network of N agents, each of which has a local uncertain estimate of θ determined by probability distributions of the same form but with local hyperparameters ω_i, $i \in \mathcal{N}$, which include some common hyperparameters ω^-.*

(a) *The Bayesian fused distribution on Θ across the team, denoted as $f_{\Theta|\Omega_{1:N}}(\theta|\omega_{1:N})$, can be determined as*

$$f_{\Theta|\Omega_{1:N}}(\theta|\omega_{1:N}) = f_{\Theta|\Omega}(\theta|\omega_{fused}), \tag{12}$$

where $\omega_{fused} = \omega^- + \sum_{i\in\mathcal{N}} \Delta\omega_i$.

(b) *If, in addition to its initial hyperparameters, ω_i, each agent $i \in \mathcal{N}$ takes M_i unique local measurements, defined as the set $Z_i \triangleq \{z_{i,1}, \dots, z_{i,M_i}\}$, and shares M_* global measurements, defined as the set $Z_* \triangleq \{z_{*,1}, \dots, z_{*,M_*}\}$, of the same underlying stationary stochastic process, $f_{Z|\Theta}$, then the fused distribution is of the same form as (12) but with hyperparameters determined as:*

$$\omega_{fused} = \omega^- + \sum_{z\in Z_*} h(z, f_{Z|\Theta}) + \sum_{i\in\mathcal{N}} \left(\Delta\omega_i + \sum_{z\in Z_i} h(z, f_{Z|\Theta}) \right). \tag{13}$$

Proof. (a) The proof is straightforward as a special case of Lemma 2(b) when $n = N$.

(b) Noting that the fused distribution in (12) is conjugate to the measurement likelihood, then the fused distribution with measurements $Z_i \bigcup Z_*$ can be found by simply applying (2) for each unique measurement, immediately yielding the desired result. □

Note that part (a) of the proof does not rely explicitly on a measurement model, such that the result is a valid Bayesian fusion even for consensus scenarios that are not measurement-based. Finally, if there is no shared information, then the shared hyperparameters are set to zero.

3.2. HPC algorithm

Having defined the fused distribution that the agents need to agree upon, this section derives the HPC method as a means to achieve this result. The basic premise behind the method is to use existing linear consensus results as highlighted in Section 2.2 to achieve a sum-consensus to the hyperparameters representative of the Bayesian fused distribution. The HPC algorithm consists of iterating the following three main actions.

Initialization: At the beginning of a consensus, each agent must initialize its hyperparameters according to (14) so that common information is accounted for and the proper consensus result is achieved.

$$\omega_i[0] \equiv \omega^- + (\omega_i - \omega^-)/\nu_i = \omega^- + \Delta\omega_i/\nu_i, \quad \forall i. \tag{14}$$

In (14), ν_i is the i-th entry of the consensus eigenvector ν and the resulting $\omega_i[0]$ serves as the initial information used in the consensus protocol.

Consensus protocol: After initialization, the agents all run the linear consensus protocol in (15) using the initialized values at iteration $k = 0$.

$$\begin{aligned}\omega_i[k+1] &= \omega_i[k] + \epsilon \sum_{j\in\mathcal{N}_i} (\omega_j[k] - \omega_i[k]) \\ &= \sum_{j\in\mathcal{N}} a_{ij}\omega_j[k].\end{aligned} \tag{15}$$

In (15), ϵ is the weighting constant and a_{ij} is the (i, j)th entry of the adjacency matrix.

Measurement update: If, at some iteration κ, agent i takes an independent measurement, z_i, of the underlying process $f_{Z|\Theta}$, it must then update its local hyperparameters according to

$$\omega_i[\kappa] \leftarrow \omega_i[\kappa] + h(z_i, f_{Z|\Theta})/\nu_i. \tag{16}$$

If the measurement is shared among all agents, such that $z_i = z_*$, $\forall i$, then each agent must update their hyperparameters as

$$\omega_i[\kappa] \leftarrow \omega_i[\kappa] + h(z_*, f_{Z|\Theta}). \tag{17}$$

The following theorem shows that, together with some mild assumptions on the network and consistency of representation, these three actions guarantee convergence to the Bayesian fused hyperparameters. From these hyperparameters, each agent then re-constructs its probability distribution on Θ and, subsequently, uses an agreed-upon loss function (such as the Minimum Mean–Squared Error, MMSE, or others (Gelman et al., 2004)) to evaluate the posterior distribution and select an appropriate realization of Θ. Thus, the agents are able to implicitly come to agreement on a best estimate of θ based on the Bayesian fusion of their local uncertainty distributions.

Theorem 2 (*Convergence of HPC*). *A group of N agents will come to an asymptotic agreement on the Bayesian fused distribution on an uncertain parameter, θ, in the presence of global common information and a finite number of concurrent measurements, by running the algorithm defined by (14) through (17) on the local hyperparameters, under the following assumptions:*

(i) *The connectivity graph $\mathcal{G}$ is time-invariant, strongly connected and the consensus eigenvector ν is known.*
(ii) *All agents use the consensus protocol defined in (15) and initialized in (14), with $\epsilon \in (0, 1/\max_i |\mathcal{N}_i|)$.*
(iii) *The agents' maintain uncertain local estimates of θ through $f_{\Theta|\Omega}$, the conjugate distribution to $f_{Z|\Theta}$, for which the conjugacy property is satisfied.*
(iv) *There may exist some shared information among all agents in the network, denoted by ω^-, conditioned upon which each agent's initial hyperparameters are independent of all other agent's hyperparameters.*
(v) *The agents have decided a priori upon the form of the distribution to use and upon which loss function will be used to determine the best estimate of θ.*
(vi) *There exists some consensus iteration $K < \infty$ after which no new measurements are made.*

Proof. Property (i) ensures that the known consensus eigenvector ν is strictly positive so that the initialization procedure in (14) is well-defined, and, combined with Property (ii), it also establishes the existence of a steady-state consensus value (Ren et al., 2007). Properties (iii)–(vi) imply that each agent will select the same *best* guess for θ by applying the same loss function to their local distribution, and that, from Theorem 1(b), the fused hyperparameters including a finite number of unique and shared measurements at iteration $k > K$ are given by

$$\omega_{\text{fused}}[k] = \omega^- + \sum_{i\in\mathcal{N}} \left(\Delta\omega_i + \sum_{z\in Z_i} h(z, f_{Z|\Theta}) \right) + \sum_{z\in Z_*} h(z, f_{Z|\Theta}), \tag{18}$$

where $Z_i \triangleq \{z_{i,1}, \ldots, z_{i,M_i}\}$ are the M_i total unique local measurements for each agent i, and $Z_* \triangleq \{z_{*,1}, \ldots, z_{*,M_*}\}$ are the M_* total shared global measurements taken during iterations up to and including iteration K. Therefore, if all agents can be shown to converge to the sum of the hyperparameter values as in (18), then they can each form the fused posterior distribution and evaluate it to obtain a consistent, fused estimate of θ, and the proof will be complete.

If all the agents follow (14) through (17) and make independent measurements at iteration $k \le K$, then, using a scalar set of hyperparameters to simplify the presentation and clarify the analysis, the resulting consensus update from iteration k to $k+1$ can be found as follows. First, define $\omega[k] \triangleq [\omega_1[k], \ldots, \omega_N[k]]^T$ as the state of the consensus variables at iteration k, and let $h_i[k]$ be the total hyperparameters contributed to agent i by the set of unique local measurements, $Z_i[k] \triangleq \{z_{i,1}[k], \ldots, z_{i,M_i[k]}[k]\}$, as well as the set of shared measurements, $Z_*[k] \triangleq \{z_{*,1}[k], \ldots, z_{*,M_*[k]}[k]\}$, taken during iteration k. Using (16) and (17), this becomes:

$$h_i[k] \triangleq \sum_{z \in Z_i[k]} \frac{h(z, f_{Z|\Theta})}{\nu_i} + \sum_{z \in Z_*[k]} h(z, f_{Z|\Theta}).$$

Next, defining $\mathbf{h}[k] \triangleq [h_1[k], \ldots, h_N[k]]^T$, the consensus update across all agents from iteration k to $k+1$ is given simply by $\omega[k+1] = A(\omega[k] + \mathbf{h}[k])$. As $k \to \infty$, the hyperparameters for each agent are given by

$$\begin{aligned}
\lim_{k\to\infty} \omega[k] &= \lim_{k\to\infty} A^k \omega[0] + A^{k-K-1} \sum_{\kappa=0}^{K} A^{K-\kappa+1} \mathbf{h}[\kappa] \\
&= \mathbf{1}\nu^T \omega[0] + \mathbf{1}\nu^T \sum_{\kappa=0}^{K} A^{K-k+1} \mathbf{h}[\kappa] \\
&= \mathbf{1} \sum_{i\in\mathcal{N}} \nu_i (\omega^- + \Delta\omega_i/\nu_i) + \mathbf{1} \sum_{\kappa=0}^{K} \nu^T \mathbf{h}[\kappa] \\
&= \mathbf{1} \sum_{i\in\mathcal{N}} \nu_i \Bigg((\omega^- + \Delta\omega_i/\nu_i) \\
&\quad + \sum_{\kappa=0}^{K} \Bigg(\sum_{z\in Z_i[k]} h(z, f_{Z|\Theta})/\nu_i + \sum_{z\in Z_*[k]} h(z, f_{Z|\Theta}) \Bigg) \Bigg) \\
&= \mathbf{1} \Bigg(\omega^- + \sum_{i\in\mathcal{N}} \Bigg(\Delta\omega_i + \sum_{z\in Z_i} h(z, f_{Z|\Theta}) \Bigg) \\
&\quad + \sum_{z\in Z_*} h(z, f_{Z|\Theta}) \Bigg),
\end{aligned}$$

with $Z_i \triangleq \cup_{k=0}^{K} Z_i[k]$, $Z_* \triangleq \cup_{k=0}^{K} Z_*[k]$, and $\mathbf{1}$ is the vector of all ones. Some steps were facilitated by invoking $\nu^T A^m = \nu^T,\ \forall m \in \mathbb{N}$ and $\sum_{i=1}^{N} \nu_i = 1$. Thus, it follows immediately that the agents converge to (18), and, implicitly, to the fused distribution itself. □

Remark 1 (*With Concurrent Measurements*)**.** In the presence of concurrent measurements during consensus, note the following observations:

(1) While no measurements are made, the agents will asymptotically converge to the current fused hyperparameters. Whenever a measurement is introduced, however, a step change in the local and fused hyperparameters occurs. In practice, the step change in the local estimate of θ for each additional measurement often decreases, such that, even though the hyperparameters may not have converged, the agents may still converge to a consistent estimate of θ.

(2) The use of a *consensus* protocol (as opposed to a traditional data fusion messaging scheme) permits the agents to deal with common information only once — specifically, when it is incorporated into the consensus either through the initialization or the specified hyperparameter updates. The trade-off for ignoring the distribution of common information once agents have begun sharing is that we need to know ν for the network so that we can ensure the desired outcome.

Additionally, it is worthwhile to note the following: the time invariance in Property (i) in Theorem 2 is a technical assumption that is included to simplify the proof, but not explicitly required (see Fraser (2009, Proposition 2.3.1)). Knowledge of ν is not a limit unique to this method since it is often implicitly achieved by assuming a balanced network and any analysis of limiting consensus values on an unbalanced network requires knowledge of ν or an equivalent. Finally, the derivation of the HPC algorithm herein has focused on the utilization of the baseline linear consensus protocol summarized in Section 2.2; however, a number of alternative average consensus architectures (for example, the dynamic average consensus in Zhu & Martínez, 2010) could be utilized within the HPC framework.

4. Numerical examples

This section elucidates the preceding derivation of the HPC method through two demonstrative examples.

4.1. Example #1: mean of a normal distribution

The hyperparameter consensus method can perform a Bayesian estimation update to find the true mean, x, of a stationary Gaussian process of known covariance by maintaining a Gaussian prior on x and a Gaussian likelihood distribution. The hyperparameters in this scenario are the parameters of the information form of the Kalman filter,

$$y = \Sigma^{-1}\hat{x}, \qquad Y = \Sigma^{-1},$$

where Σ is the covariance matrix of the prior distribution and $\hat{x}$ is its mean (and thus the best estimate of x). The Bayesian hyperparameter measurement updates are $y \leftarrow y + \mathbf{i}$ and $Y \leftarrow Y + \mathbf{I}$, where $\mathbf{i}$ denotes the modified measurement vector, and $\mathbf{I}$ is the inverse of the covariance matrix of the likelihood distribution (Grime et al., 1992; Maybeck, 1979; Xiao et al., 2005). For example, for the linear sensing model, $z = x + v$, $v \sim \mathcal{N}(0, R)$ with R denoting the (known) covariance of the process being measured, the modified measurement vector is a function of the observation, $\mathbf{i} = R^{-1}z$, and the inverse measurement noise covariance becomes $\mathbf{I} = R^{-1}$. Since the hyperparameter updates are now of the additive form, HPC can be applied to each hyperparameter concurrently, with the unique local information defined as: $\Delta\omega_i = [\Delta\mathbf{i}_i,\ \Delta\mathbf{I}_i]^T = [y_i - y^-,\ Y_i - Y^-]^T$. Theorem 2 then states that the hyperparameter consensus will converge to the desired Bayesian fusion of the hyperparameters:

$$\lim_{k\to\infty} y_i[k] = y^- + \sum_{j\in\mathcal{N}} \Delta\mathbf{i}_j, \qquad \lim_{k\to\infty} Y_i[k] = Y^- + \sum_{j\in\mathcal{N}} \Delta\mathrm{I}_j.$$

In case each agent makes no measurements and has entirely independent information, the fused covariance and mean can be found from the priors:

$$\Sigma_{\text{fused}} = \left[\sum_{i\in\mathcal{N}} Y_i\right]^{-1}, \qquad \hat{x}_{\text{fused}} = \Sigma_{\text{fused}} \left[\sum_{i\in\mathcal{N}} Y_i \hat{x}_i\right].$$

In this case, the local state and covariance estimates can be retrieved and verified as the true fused parameters:

$$\Sigma_{i_{\text{fused}}} = Y_{i_{\text{fused}}}^{-1} = \left[\sum_{j\in\mathcal{N}} Y_j\right]^{-1} = \Sigma_{\text{fused}}$$

$$\hat{x}_{i_{\text{fused}}} = Y_{i_{\text{fused}}}^{-1} y_{i_{\text{fused}}} = \Sigma_{\text{fused}}\left[\sum_{j\in\mathcal{N}} y_j\right] = \hat{x}_{\text{fused}}.$$

It can be shown that previous Kalman consensus approaches lead to zero (Ren et al., 2005) or biased (Alighanbari & How, 2008) estimates of the covariance (Fraser, 2009; Fraser et al., 2009). Thus, for estimation of the mean of a stationary Gaussian process, the HPC algorithm expands upon the results obtainable using previous Kalman consensus algorithms by allowing the agents to not only converge to the unbiased fused mean, but also to maintain a representative level of confidence in their fused estimate after consensus through the proper fused covariance.

4.2. Example #2: Poisson arrival rate

This example will focus on the use of a gamma prior on the arrival rate parameter, λ, of a Poisson distribution. The gamma and Poisson distributions, respectively, are:

$$f_{\Lambda|A,B}(\lambda|\alpha,\beta) = \beta^{\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}/\Gamma(\alpha),$$

$$f_{X|\Lambda,T}(x|\lambda,t) = (\lambda t)^x e^{-\lambda t}/x!, \quad x \in \mathbb{Z}_{\geq 0}.$$

The hyperparameters in this situation are the scalars α and β, which can be shown to be updated after a measurement as $\alpha \leftarrow \alpha + x$, $\beta \leftarrow \beta + t$. In other words, α and β represent the number of events occurred and the time interval during which that many events have happened. Using the MMSE metric, the best estimate of λ is the mean of the gamma distribution, $\lambda_{\text{MMSE}} = \frac{\alpha}{\beta}$, and the fused MMSE estimate is found as:

$$\lambda_{\text{fused}} = \sum_{i\in\mathcal{N}} \alpha_i \Big/ \sum_{i\in\mathcal{N}} \beta_i.$$

The following results are for five agents agreeing over a known, unbalanced network with a consensus eigenvector of $\nu = [\frac{1}{16}\ \frac{1}{16}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{2}]^T$, where each agent initial has different values of hyperparameters.

Fig. 1 shows the MMSE parameter trajectory during consensus with independent measurements made by two agents at iteration $k = 5$, and with the fused MMSE estimate denoted by the dotted red line. The difference between the "Initial" and "Weighted" estimate is the initialization in (14), which affects each agent differently due to the presence of shared information and the fact that the network is unbalanced. Once the consensus begins, the agents quickly begin to converge to the fused estimate of $\lambda = 1$. In this scenario, two agents take independent measurements during the transient period of the initial consensus, at which point the agents immediately begin to converge to the new fused estimate.

4.2.1. Sensitivity to unknown consensus eigenvector

HPC works best on known networks so that the agents can weight their initial conditions according to the value of ν. However, even when the network is unknown and the initial conditions cannot be weighted properly, the HPC method with $\nu_i = 1/N$ performs better than pure linear consensus on the parameter itself (i.e., running (3) on $\lambda_{\text{MMSE}_i} = \alpha_i/\beta_i$). A comparison is made between the expected percent error in the parameter estimate for the hyperparameter and parameter consensus methods, where the errors are defined as $\hat{e} = |\hat{\lambda}/\lambda_{\text{fused}} - 1|$ and $\bar{e} = |\bar{\lambda}/\lambda_{\text{fused}} - 1|$, respectively, and where $\hat{\lambda}$ is the steady-state HPC estimate and $\bar{\lambda}$ is the parameter consensus estimate. Fig. 2 shows the amount by which the expected parameter consensus error is greater than the expected HPC error for different network sizes as a function of the degree of *biasedness* of the network (Fraser, 2009), where a network has a bias of 0 if it is balanced and a bias of 1 if $\nu = [0\ \ 0\ \ \ldots\ \ 0\ \ 1]^T$ (i.e., if the network is a directed spanning tree and not strongly connected). The plotted results are obtained by sampling 200,000 Monte Carlo simulations for each data point, where the initial α for each agent is sampled uniformly between 5 and 10, inclusive, and initial β is a sum of α samples from an exponential distribution with $\lambda = 1$. It shows that, for the given initial conditions and for all values of network bias, it is expected that the HPC method will achieve an MMSE parameter estimate that is closer to the proper fused estimate than linear consensus on the parameter, even when the network is unknown.

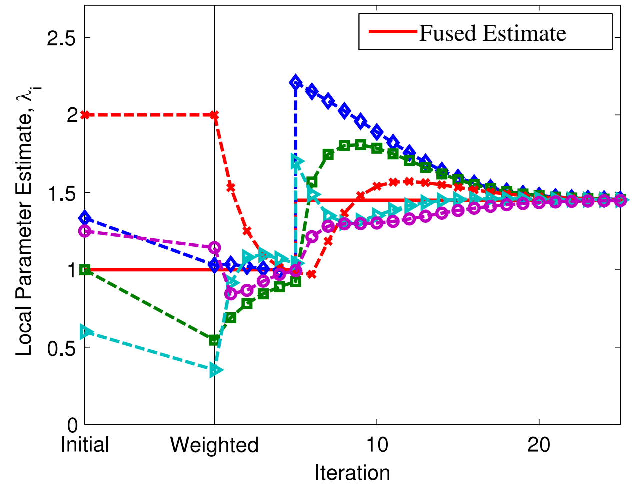

Fig. 1. Local parameter estimates using hyperparameter consensus on an unbalanced network with measurements.

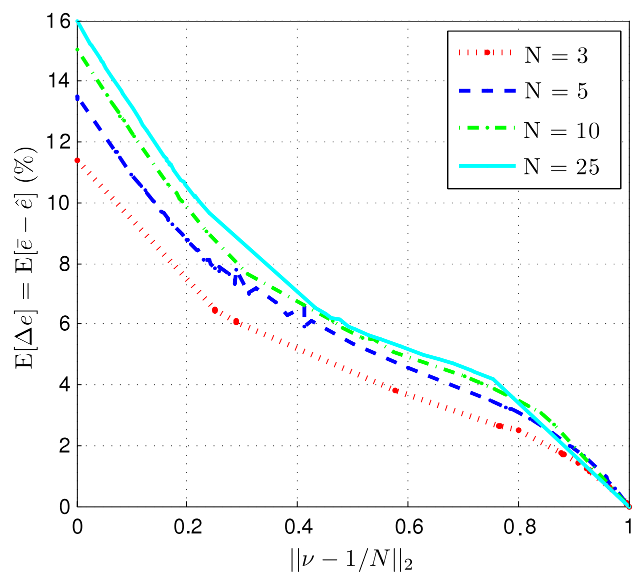

Fig. 2. Expected difference in errors ($E[\bar{e} - \hat{e}]$) over an unknown network.

5. Conclusions

This paper presented the hyperparameter consensus method that is proven to allow a network of agents to come to an unbiased distributed agreement to the Bayesian fusion of their local parameter estimates defined by possibly non-Gaussian uncertainties. The primary innovation of this method is the use of the hyperparameters of local distributions that effectively transforms the Bayesian fusion problem into a standard sum consensus problem. Numerical examples demonstrated that the proposed method facilitates expedited learning ability over the network.

Acknowledgments

This work was sponsored (in part) by the AFOSR and USAF under grant FA9550-08-1-0086. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the US Government.

This work was also in part supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0025484).

References

Alighanbari, M., & How, J. P. (2008). Unbiased Kalman consensus algorithm. *Journal of Aerospace Computing, Information, and Communication*, *5*(9), 298–311.

Beard, R., & Stepanyan, V. (2003). Information consensus in distributed multiple vehicle coordinated control. In *IEEE conference on decision and control* (pp. 2029–2034).

Cortés, J. (2008). Distributed algorithms for reaching consensus on general functions. *Automatica*, *44*(3), 726–737.

Fax, J., & Murray, R. (2004). Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, *49*(9), 1465–1476.

Fraser, C. S. R. (2009). Reaching consensus with uncertainty on a network. Master's thesis. MIT Department of Aeronautics and Astronautics.

Fraser, C. S. R., Bertuccelli, L. F., & How, J. P. (2009). Reaching consensus with imprecise probabilities over a network. In *AIAA conference on guidance, navigation and control* (AIAA–2009–5655).

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis* (2nd ed.). Chapman and Hall.

Grime, S., & Durrant-Whyte, H. (1994). Data fusion in decentralized sensor networks. *Control Engineering Practice*, *2*(5), 849–863.

Grime, S., Durrant-Whyte, H., & Ho, P. (1992). Communication in decentralized data-fusion systems. In *American control conference* (pp. 3299–3303).

Hatano, Y., & Mesbahi, M. (2005). Agreement over random networks. *IEEE Transactions on Automatic Control*, *50*(11), 1867–1872.

Jadbabaie, A., Lin, J., & Morse, A. S. (2003). Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, *48*(6), 988–1001.

Lin, J., Morse, A., & Anderson, B. (2003). The multi-agent rendezvous problem. In *IEEE conference on decision and control* (pp. 1508–1513).

Makarenko, A., & Durrant-Whyte, H. (2006). Decentralized Bayesian algorithms for active sensor networks. In *International conference on information fusion*, 7 (4) (pp. 418–433).

Maybeck, P. S. (1979). *Estimation and control*: *Vol. 1*. *Stochastic models*. Academic Press.

Olfati-Saber, R. (2006). *Network embedded sensing and control*: *Vol. 331*. *Distributed Kalman filtering and sensor fusion in sensor networks* (pp. 157–167). Springer-Verlag.

Olfati-Saber, R., Fax, J., & Murray, R. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, *95*(1), 215–233.

Olfati-Saber, R., & Shamma, J. (2005). Consensus filters for sensor networks and distributed sensor fusion. In *IEEE conference on decision and control and European control conference* (pp. 6698–6703).

Ren, W. (2006). Consensus based formation control strategies for multi-vehicle systems. In *American control conference* (pp. 6–12).

Ren, W., Beard, R. W., & Atkins, E. M. (2007). Information consensus in multivehicle cooperative control. *IEEE Control System Magazine*, *27*(2), 71–82.

Ren, W., Beard, R. W., & Kingston, D. B. (2005). Multi-agent Kalman consensus with relative uncertainty. In *American control conference* (pp. 1865–1870).

Ren, W., Beard, R. W., & McLain, T. W. (2005). Coordination variables and consensus building in multiple vehicle systems. In *Lecture notes in control and information sciences*: *Vol. 309* (pp. 171–188).

Tahbaz-Salehi, A., & Jadbabaie, A. (2008). A necessary and sufficient condition for consensus over random networks. *IEEE Transactions on Automatic Control*, *53*(3), 791–795.

Waltz, E., & Llinas, J. (1990). *Multisensor data fusion*. London: Artech House Boston.

Wu, C. W. (2006). Synchronization and convergence of linear dynamics in random directed networks. *IEEE Transactions on Automatic Control*, *51*(7), 1207–1210.

Xiao, L., Boyd, S., & Lall, S. (2005). A scheme for robust distributed sensor fusion based on average consensus. In *International symposium on information processing in sensor networks* (pp. 63–70).

Zhu, M., & Martínez, S. (2010). Discrete-time dynamic average consensus. *Automatica*, *46*(2), 322–329.

Cameron S.R. Fraser received his Master of Science from the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology in 2009. His research interests include distributed estimation and decentralized decision making algorithms as applied to networks of unmanned vehicles. Since graduating, he has worked with the Canadian Centre for Unmanned Vehicle Systems to examine methodologies for the regulation and standardization of current and future unmanned aircraft systems, and is currently working on control law design in the fly-by-wire group at Bombardier Aerospace in Montréal.

Luca F. Bertuccelli joined the United Technologies Research Center in East Hartford, CT, as a Senior Research Engineer in August 2011. He obtained his B.S. in Aeronautical and Astronautical Engineering at Purdue University in 2002, his M.S. and Ph.D. in Aeronautics and Astronautics at MIT in 2004 and 2008 respectively, and was previously a Postdoctoral associate at MIT in the area of human supervisory control. His Ph.D. thesis addressed robustness to modeling errors in Markov Decision Processes with applications to UAV systems and presented a new sampling method for solving robust MDPs for aerospace applications. His current research interests are in robust planning, integrated human–machine systems, and distributed decision making.

Han-Lim Choi is an Assistant Professor of Aerospace Engineering at KAIST (Korea Advanced Institute of Science and Technology). He received his B.S. and M.S. degrees in Aerospace Engineering from KAIST, Daejeon, Korea, in 2000 and 2002, respectively, and his Ph.D. degree in Aeronautics and Astronautics from Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2008. He then studied for one and a half years at MIT as a postdoctoral associate until he joined KAIST in 2010. His current research interests include estimation and control for sensor networks and decision making for multi-gent systems. He (together with Dr. Jonathan P. How) is the recipient of 2011 Automatica Applications Paper Prize, is a member of the American Institute of Aeronautics and Astronautics (AIAA), and a member of the Institute of Electrical and Electronics Engineers (IEEE).

Jonathan P. How is the Richard C. Maclaurin Professor of Aeronautics and Astronautics at the Massachusetts Institute of Technology (MIT). He received a B.A.Sc. from the University of Toronto in 1987 and his S.M. and Ph.D. in Aeronautics and Astronautics from MIT in 1990 and 1993, respectively. He then studied for two years at MIT as a postdoctoral associate for the Middeck Active Control Experiment (MACE), which flew on-board the Space Shuttle Endeavour in March 1995. Prior to joining MIT in 2000, he was an Assistant Professor in the Department of Aeronautics and Astronautics at Stanford University. Current research interests include robust coordination and control of autonomous vehicles in dynamic uncertain environments. He was the recipient of the 2002 Institute of Navigation Burka Award, is an Associate Fellow of AIAA, and a senior member of IEEE.